CENTRE DE RECHERCHE
SUR LES TRANSPORTS
(C.R.T.)
CENTRE FOR RESEARCH
ON TRANSPORTATION

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE

CENTRE DE RECHERCHE SUR LES TRANSPORTS
UNIVERSITÉ DE MONTRÉAL
C.P. 6128, SUCCURSALE A
MONTRÉAL
CANADA H3C 3J7

TÉLÉPHONE : (514) 343-7575
EMAIL : crt@crt.umontreal.ca
TÉLÉCOPIEUR (FAX) : (514) 343-7121

CRT-835

Dynamic and Stochastic Models for the Allocation of Empty Containers

by

Teodor Gabriel Crainic[1, 2]
Michel Gendreau[1]
Pierre Dejax[3]

[1] Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, Canada H3C 3J7.
[2] Département des sciences administratives, Université du Québec à Montréal, Montréal, Canada.
[3] Laboratoire Productique et Logistisque, Ecole Centrale Paris, Chatenay-Malabry, France.

Centre de recherche sur les transports — Publication #835

October 1992

Dépôt légal – Bibliothèque nationale du Québec,
Bibliothèque nationale du Canada, 1992

93-01830/

ABSTRACT

The empty container allocation problem occurs in the context of the management of the land distribution and transportation operations of international maritime shipping companies. It involves dispatching empty containers of various types in response to requests by export customers and repositioning other containers to storage depots or ports in anticipation of future demands. We describe the problem and identify its basic structure and main characteristics. We then introduce two dynamic deterministic formulations, for the single and multicommodity cases, which offer a general modeling framework for this class of problems, and which account for its specific characteristics: the space and time dependency of events, substitutions among container types, relationships with partner companies, imports and exports, massive equilibration flows, etc. Finally, we provide a mathematical formulation for handling, in the single commodity case, the uncertainty of demand and supply data, characteristic of container allocation and distribution problems. Various modeling choices, data requirements, and algorithmic considerations related to the implementation of the models are also discussed.

KEYWORDS Dynamics and Stochastic Models;
Multicommodity Network Models with Gains;
Vehicle Allocation; Fleet Management;
Land Container Transportation.

RESUME

Le problème d'affectation de conteneurs vides se pose dans le contexte de la gestion des opérations de distribution et de transport terrestre des compagnies maritimes internationales. Il s'agit d'envoyer des conteneurs vides de différents types aux clients exportateurs qui en font la demande ainsi que d'en repositionner d'autres, à des dépôts intérieurs ou portuaires, afin d'être en mesure de satisfaire les demandes futures. Nous décrivons le problème, ses élements fondamentaux et ses principales caractéristiques: les dépendances spatio-temporelles des offres et des demandes, les substitutions entre containers de types différents, les échanges avec d'autres companies, les importations et exportations, l'existence de flots massifs d'équilibrage inter-dépôts, etc. Deux formulations déterministes, une pour le cas à un produit et l'autre traitant le cas de plusieurs types de conteneurs, sont introduites, ce qui constitue un cadre de modélisation complet pour ce problème. Nous présentons également une formulation mathématique permettant de traiter l'incertitude caractérisant les offres et les demandes. Nous discutons également de certains choix de modélisation, des exigences de disponibilité de données, et des perspectives algorithmiques, en vue d'applications pratiques des modèles.

MOTS CLES Modèles dynamiques et stochastiques;
Modèles de flots dans les réseaux avec gains;
Affectation de véhicules;
Gestion de flotte; Transport terrestre de conteneurs.

INTRODUCTION

Since the seventies, freight transportation activities have seen many fundamental changes both in terms of technology and of organization. In particular, one must note the increasing use of multimodal transportation routes and of containerized transportation practices. This is particularly true in the area of international maritime shipping, where the maritime carrier normally owns the containers and insures their multimodal transportation from the initial origin to their final consignee.

We are particularly interested in the planning of the land transportation part of the international maritime shipping of containers. This is an extremely complex activity, especially if one aims to simultaneously optimize the cost and service aspects of the company's operations in a competitive environment. A fundamental component of this problem is concerned with the allocation of empty containers to customers, in preparation of future loaded export movements, and the repositioning of other empty containers to storage depots or ports to be used to fulfill next periods' orders (both expected and forecast).

To our knowledge, there does not exist any other model specifically aimed at the empty container allocation problem that we study. In a recent survey of the literature on fleet management models in freight transportation, Dejax and Crainic (1987) noted that relatively little effort was directed toward developing models aimed specifically toward container transportation issues, with the bulk of the work in the area being dedicated to the maritime aspects of the problem. Thus, Dejax and Crainic quote a few authors who addressed the problem of the empty container distribution as that of allocating containers available in surplus at a terminal, from earlier loaded shipments, to demanding terminals (e.g. a customer's warehouse, a rail yard, or a port), in preparation for subsequent loaded movements. Antonisse (1988) is a recent example of such efforts, for the worldwide maritime container fleet management problem.

The state of the art described by Dejax and Crainic may be explained, at least partially, by the combined effect of the relative complexity of the container land transportation issues and the opinion, prevalent until very recently in the industry, that the impact of land operations on the company's overall performance was marginal. This perception is, however, gradually evolving and the land portion of the container journey is receiving increasing attention.

This is especially true in today's increasingly competitive environment where growing customer demand is associated to the search for cost efficient high quality service operations. This view is similar to those of, among others, Jarke (1982), who reviews container management problems and who insists on the necessity for efficient information and decision-support systems for container transportation logistics, and Antush (1989), who emphasizes the need for developing new information systems and planning methodologies for container transportation.

The goal of this paper is to propose models for empty container allocation in a land distribution and transportation system, which address the complexity of the specific problems which arise in this area, particularly the space and time dependency of events and the uncertainty of supply and demand data, plus specific operational characteristics such as substitutions among container types, relationships with partner companies, imports and exports, massive equilibration flows, etc. The focus of the paper is on modeling issues, its main contribution being threefold: (i) identification of the basic elements and structure of the container distribution problem; (ii) development of a multicommodity model that integrates substitutions of container types; (iii) providing a mathematical formulation for handling the dynamic aspects and the uncertainties characteristic of container allocation and distribution problems.

The paper is organized as follows. In the first section, we give a brief description of the issues related to planning container distribution and transportation systems, as well as of the general methodological framework we propose to address them. The land container allocation problem, its main components and characteristics are described in Section 2. Section 3 is dedicated to the development of the mathematical and network formulations of the deterministic, single commodity version of the model, while the multicommodity version of the model is presented in Section 4. The dynamic stochastic single commodity formulation is treated in Section 5. In Section 6 we compare our modeling framework with the work of other authors, while Section 7 is dedicated to considerations related to the practical implementation of the models. Our conclusions indicate future research directions in this area.

1. PROBLEM CONTEXT

Our research is motivated by the analysis of typical land transportation problems encountered by international shipping companies that operate maritime and land networks of considerable, even continental, scale. Crainic, Gendreau and Dejax (1990) describe the various operations and planning issues involved, and present the methodological framework that we propose in order to address them. To facilitate the understanding of the allocation model, this section is dedicated to a brief overview of these issues and design principles.

Arriving ships carry containers, which generally come in two sizes and about ten major types each, loaded with imported goods, as well as empty containers returning from previous exports. Loaded containers must further be transported to their final destination, usually an industrial firm or a distribution center located somewhere in the continent. Ground transportation may be performed by using rail, truck, barges and mixed modes. This latter option usually consists of an initial railway shipment to an inland railway terminal, followed by a final movement by truck to the customer. The empty containers that are imported may either stay at the port for a while, or be immediately dispatched wherever they are needed for subsequent exports.

Once unloaded, empty containers have to be picked up at the customer's site and moved away. They may either return to the port of origin, or may be transported to any other depot: port or inland terminal. This return shipment may, but more often than otherwise will not, be made using the same vehicle (truck, railway car, ...) that has been used for the transportation from the port to the customer. Similarly, exporting firms need empty containers to be delivered to them, either from the future export port, or from any other location. Due to complex organizational considerations, very few (Dejax, Crainic and Delorme, 1988, report some 3% for the case they studied) empty container movements are made directly from import to export customers. After loading, containers are transported to the export port and are loaded on ships together with empty containers sent abroad to cope with the worldwide imbalance in the supply/demand of certain container types.

Note the importance of empty container movements. First, every commercial (profitable) movement of a loaded container generates almost automatically a non-profitable empty container movement: once loaded containers are delivered to a customer's site and are unloaded,

they normally have to be moved away empty, while empty containers have to be delivered to exporting customers. Moreover, significant regional imbalances between imports and exports within the area served by the company (their amplitude may vary according to the container type and the geographical subdivisions considered), as well as opportunities for cost reductions due to mass transportation of containers over relatively long distances (by railway, for example), result in land movements of empty containers directly between depots: these are called *balancing flows*.

Thus, by their impact on the amplitude of the resource requirements and on the quality of service offered to the company's customers, empty movements play a central role in the management of the land distribution and transportation of containers. A few figures illustrate and support this point (Dejax, Crainic and Delorme, 1988). A major European container shipping company operates worldwide maritime shipping lines to and from some twenty European ports, while its land distribution and transportation network mainly covers France, Germany, Holland, Belgium and Luxembourg. In 1986, its operations involved over 300,000 land container movements, which represented an estimated total distribution and transportation cost of some U.S. $ 50,000,000. Of these, 40% were empty movements!

Several key questions have to be addressed by container shipping companies for the organization of their land distribution and transportation operations, questions which clearly belong to several planning levels: strategic (e.g. the selection of inland terminals and the determination of their logistic zones), tactical (e.g. the evaluation of balancing flows) and operational (e.g. the dynamic allocation of containers). They cannot all be represented by a single model. Moreover, they do not require the same data, nor the same personnel and resources for their planning and execution. Hierarchical relations exist among them and are reflected in the flow of information and decisions. Yet, the modeling effort is significantly facilitated by the fact that container shipping companies do not often need to build expensive inland terminals: they may in fact use already existing terminals, such as rail yards, for a cost much lower than the cost of building their own facilities. This allows them to periodically re-evaluate the decisions concerning the use of inland terminals and the corresponding allocation of customers to depots. An integrated, multilevel methodology, which reflects the observed hierarchy in the decision process and the flow of information, may then be designed in order to attempt to answer these questions.

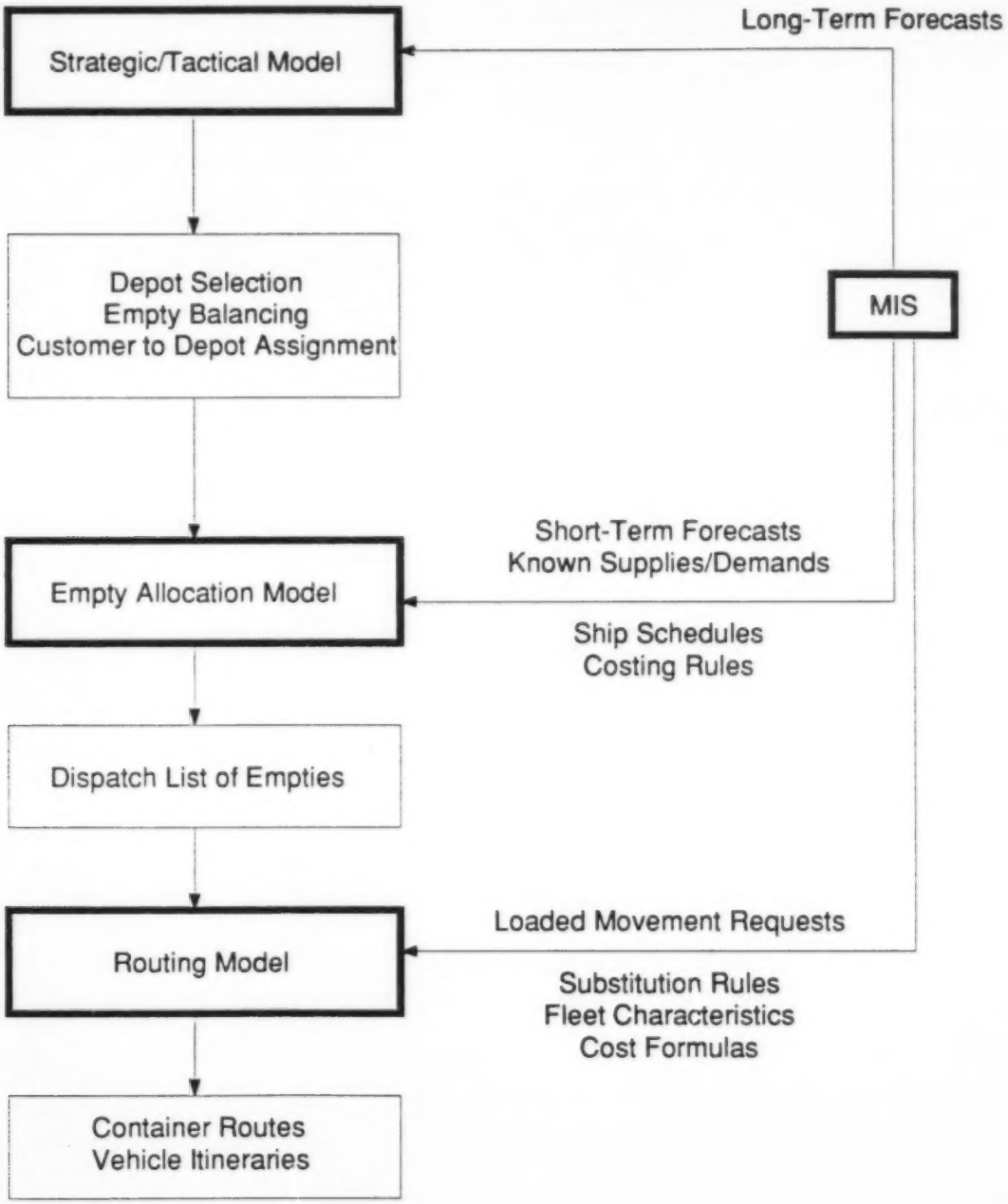

Figure 1. Outline of the Overall Planning Approach

Figure 1 displays a schematic representation of this approach. The topmost level is concerned with the strategic/tactical planning of operations. The model is formulated as a multimode, multicommodity location/distribution problem with interdepot balancing requirements (Crainic, Dejax and Delorme, 1989). Several specifically tailored algorithms (see Crainic, Gendreau and Dejax, 1990, for references) have been developed, which efficiently make full use of the particular characteristics of the formulation. The main output of this phase is the *strategic/tactical transportation plan*, which specifies the network configuration for the current planning period: the depots that are to be used, the allocation of customers and customer zones to depots, and the main interdepot balancing flows. The allocation of customers to depots is specified by container type and direction of movement, from customer to depot and from depot to customer. In general, a customer may be assigned to several depots, for each type of container and each direction of movement. Then, the strategic/tactical plan indicates the relative importance (proportions) of these flows.

Balancing traffic is also indicated by container type and direction of movement. Generally, however, these results do not represent decisions to be directly carried out in actual operations. Rather, being based upon known customer agreements and medium term demand and supply forecasts, they act as indicators of the magnitude of the balancing operations required over the following periods, and may be used to negotiate transportation tariffs (with rail companies, generally) for regular, consolidated movements of empty containers (e.g., unit trains). These agreements will then constrain the day to day operations.

The two other models correspond to the level of the operational planning of the company's activities. At this level, one has to make sure that demand is satisfied and that the most effective routes and means of transportation are selected and used, in order to achieve the best possible economic and service performance in a dynamic and stochastic environment. Ideally, one would like to develop a single mathematical model to optimize the short-term land operations of the company, in order to fully account for the interactions between the various decisions to be made. Given the intrinsic complexity of the problem at hand and the current state of the art in O.R., this is not feasible. This explains our decision to break down this large problem, and to develop two separate optimization models.

The allocation model aims to determine the "best" distribution of empty containers that satisfies both known and forecast customer demands. To allow for the high level of

uncertainty inherent in the real-time operation of a transportation system and for the impact of current decisions in the future, it should be a stochastic, dynamic model covering a planning horizon of one to two weeks. The objective of this paper is to present formulations for the allocation model.

The routing model strives to minimize the overall transportation cost of the loaded and empty containers from their origin to their destination. This is a complex and difficult problem, since it involves not only the determination of the "best" (multimodal) container itineraries, but also the construction of the routes for the vehicles required to transport them (for one or several modes). It is worthy to note that, since the container shipping company does not own, in any significant measure, the required land transportation means, substantial savings can be achieved by using a routing model. This is particularly true for truck transportation where most savings are achieved by generating multi-stop "circuit" routes (and by suitably matching requested container movements to these routes) for which tariff reductions may be negotiated. The output of the routing model is a list of movement orders which completely describe the loaded and empty movements to be executed during the next period (Dejax et al., 1992).

As illustrated in Figure 1, the use of these models requires that a wide gamut of information be retrieved from the company's information system. In particular, data on demand is to be given, including long and short-term forecasts, as well as known supplies and demands for the forthcoming periods. Equally essential are the rules that specify how transportation means are chosen and priced, plus the company's own accounting criteria concerning the various costs related to owning and using containers. In the following sections, we further discuss these information needs in relation to the requirements of the allocation model.

2. PROBLEM DESCRIPTION

Several decisions are interwoven into the empty container allocation process. These decisions are mainly concerned with *when* to initiate a transportation operation, as well as with *from where* to originate, or *where* to terminate, the operation. (At this planning stage, *how* to perform the movement – mode, route, etc. – is usually implicit in the spatial and temporal definition of the movement.) Besides these "movement" decisions, one is also concerned with operations that affect the availability of containers for allocation: stock levels at depots, substitutions among containers of various types, additional resources (e.g. new or borrowed containers), etc. These decisions require a great deal of information, including a variety of considerations regarding the company's distribution and transportation system, the levels of supplies and demands, the company's operating policies and relations with its environment (competition, partner shipping companies, official rules and regulations, agreements with carriers, etc.). Here, we describe the main decisions, system components and data considerations which make up the empty container allocation process. The emphasis is on the elements which most strongly characterize (and increase the complexity of) the problem, and hence which have significantly influenced the design of the models.

The *physical* network which corresponds to the company's distribution and transportation system is made up of ports, depots and customers, as well as of the transportation links connecting these elements. Such a network is very schematically drawn in Figure 2.

In a land container distribution and transportation system, *ports* play a special role. They are the main entry and exit points of the system, since most of the exchanges between the land transportation system and the international shipping network which contains it take place in ports. Furthermore, container shipping companies are primarily maritime industries; hence, they operate significant parts of their land network from ports, where they also maintain important stocks of empty containers. Consequently, we have decided to specifically include ports in the formulation.

"Inland" *depots* complete the backbone of the land distribution system. Physically, these depots often are located in facilities offered by other transportation modes (e.g. rail yards), the container shipping company only renting the space it needs. Consequently, the company may adjust its network configuration and operations to the demand variations. Yet,

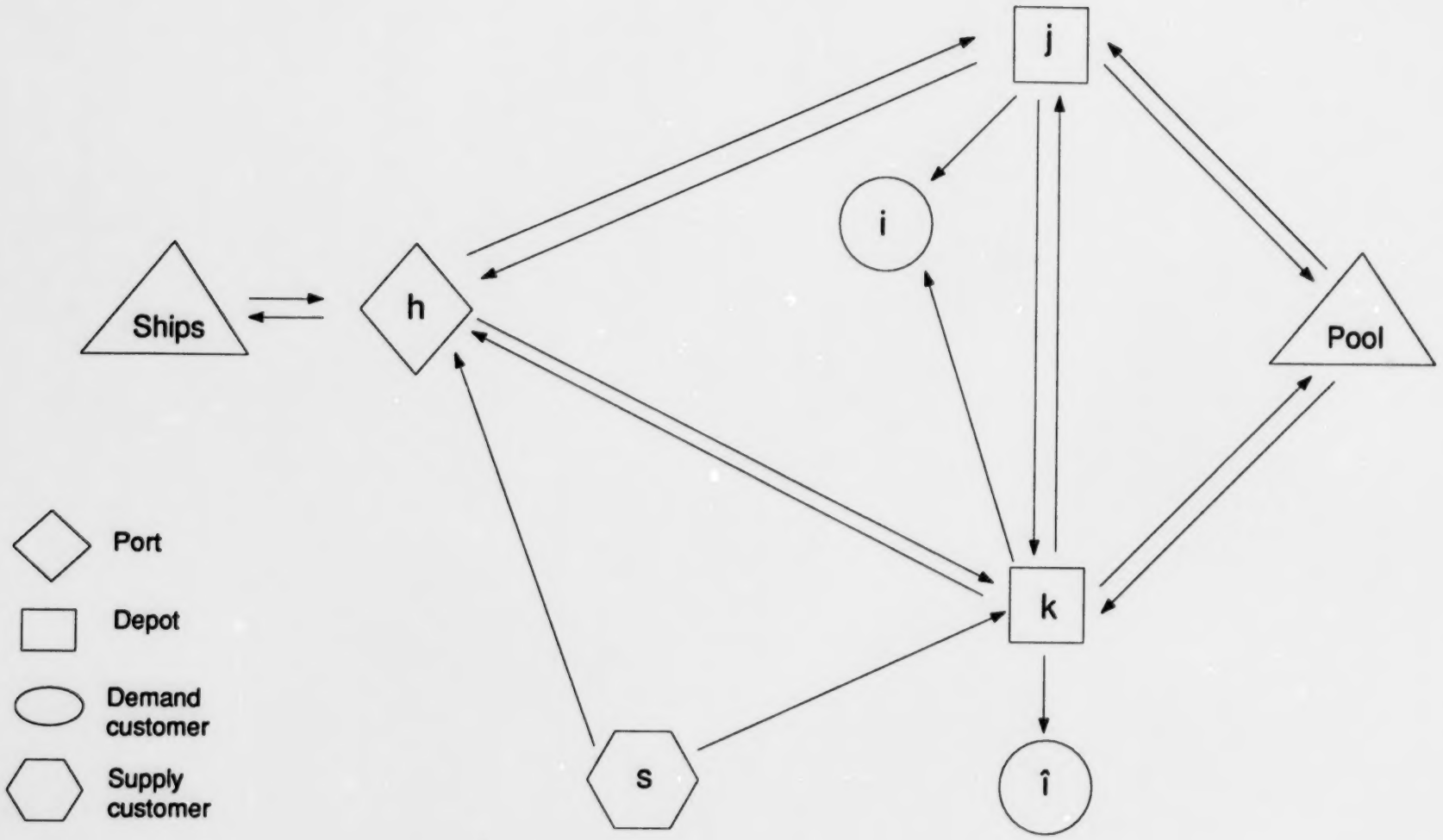

Figure 2. Possible Container Movements in a Land Transportation System

for a given planning horizon, the port and depot network configuration (depots j, k and port h of Figure 2) is fixed.

Depots, both inland and in ports, serve as primary storage and transit locations for empty containers (they may also serve as intermodal transfer points for multimodal transportation of loaded containers; this aspect is, however, quite beyond the scope of this paper): stocks of empty containers are built, demands for empties are satisfied by allocating containers out of these facilities, empty containers available at customers' sites are shipped to depots, balancing movements are initiated from there. The warehousing function and decisions are rather straightforward. In order to describe the movement decisions, however, we need first to introduce the *customers* of the system.

We distinguish between *demand* and *supply* customers (although some may be both). A demand customer requires that empty containers be delivered at its site, while a supply customer has empty containers, from a previous loaded import shipment, that have to be taken away. Over a given planning horizon, each customer may generate several supply and demand requests, each of these being characterized by the number of containers requested, their type, temporal conditions (e.g., the latest date of delivery), etc.

The location of a customer's site is known. For each demand customer, the strategic/tactical plan has also determined the depots that may satisfy its requests for each type of container. In Figure 2, for example, demand customer i may be served from both depots j and k, while the requests of customer $\hat{i}$ may only be satisfied out of depot k. For each request of a demand customer, one also generally knows the *delivery window* within which containers have to be delivered. The delivery windows are mainly determined by the due date in ports (determined by the ship schedules) and by estimations of the time necessary to load the containers and to transport them. It may be worth noting that, at the level of planning considered herein, the definition of delivery windows implies that backorders are not allowed in the system. The allocation decision hence corresponds to determining how to satisfy each customer demand request: what quantities, from what depot and when to initiate the movement.

Supply customers may either send the empty containers to one (or several) of their assigned depot(s) (depot k for customer s of Figure 2), or back to the port that originated the

loaded movement (link (s,h) of Figure 2). This is an important issue, due to its influence on the inventory levels at depots and on the subsequent allocation decisions. Thus, the input of loaded containers into the system and the moment when these containers are again available to the company have a significant impact both on routing issues and on the future *supply* of empty containers at depots.

The *dynamic* and *stochastic* characteristics of the problem are mostly evident in the definition of the supplies and demands of the system. The *empty container supply* consists of all the containers which are made available for allocation at any given time. The supply for each type of container is made up of the stocks at depots, of the containers present at supply customer sites, and of those entering the system via arriving ships in ports. Note, however, that before being available for allocation, most containers have to be moved: from supply customers to some depot, or between depots. Furthermore, not all customer supplies become available at the same moment, even if the loaded import containers are delivered from the same ship. This is mainly due to uneven transportation and unloading times.

The *empty container demand* is indicated for each customer, for each type of container, and for each period of the planning horizon. In fact, the same customer may require containers, of the same or of different types, at several periods. Therefore, one needs to distinguish between the "physical" customer (its site) and its various demand requests. Most of these demands correspond to *identified* customers, with known sites, and can be considered explicitly in the allocation model. However, while some of these demands are completely determined ("sure" orders), those of some other customers are known with less certainty, especially as the period where the demand is expected is farther down the temporal axis. Finally, it is also known that, in each region and for each type of container, there is a nonzero probability that an unforecast demand will pop up in some future period. We associate this type of demand with the depots.

As previously described, the *strategic/tactical plan* mainly specifies the network configuration for the current planning period, and the restrictions imposed upon the interdepot balancing flows by the agreements between carriers and the company. The strategic/tactical plan may be viewed as a set of operational policies in an ideal situation. In practice, however, one has to deal with the uncertainties inherent to real operations, and has therefore to depart from this target. Hence, the exact amplitude of the interdepot balancing flows is a

decision that belongs within the scope of the allocation model.

Differences between supplies and demands of containers of given types exist not only at the national and regional levels, but also at the level of the international trade. Therefore, one notes the existence of *external movements of empties*, associated with the input/output function of ports and represented by the arrival (departure) of empty containers from (to) outside the land system. Ships perform these movements, together with the import and export of loaded containers. At the decision level we consider, one has (for all practical purposes) no possibility to modify the ships' schedules. Hence, these schedules are assumed known, as are the manifests of the ships specifying the origin, destination, type and number of each shipment of (loaded or empty) containers. Consequently, in our formulation, ship arrivals and departures correspond to temporal entry and exit points of the system, determining when certain supplies become available and the latest date when some demands may be satisfied. We do not track containers outside the land transportation system.

Consideration has also to be given to the *external affairs* activities which impact upon the availability of empty containers, such as possible exchanges with partner container shipping companies, renting of empty containers, acquisition of new ones, etc. The wreckage of containers, either definitely (when they are scrapped and go out of the system) or temporarily (an estimation of the expected repair time is then needed), falls also into this category. We schematically illustrate these possibilities by the connections between depots j and k (the link from port h is omitted for clarity's sake) and the node *pool* of the network in Figure 2. Probability distributions based on historical data may be used to represent the amplitude of these activities.

An important characteristic of the problem, and one which adds considerably to its degree of complexity, is the possibility to *substitute* one type of container for another. The container *substitution rules* are specific for each pair of container types, and are not necessarily symmetric, neither in type (type A may be substituted for type B, but the reverse is not permitted), nor in number (one type A container may be substituted for a container of type B, but it takes two of these to replace one type A). Moreover, these rules may have to be adjusted according to the loading/unloading facilities available at the customer's site (one cannot substitute a 40-foot container for two 20-foot ones if the dock is not long enough), and the specific commodity that is to be transported (one cannot substitute two 20-foot containers

for a 40-foot one if anything over twenty feet long has to be transported). The handling of substitutions is further complicated by the emergence of containers of dimensions which differ from the "traditional" forty or twenty feet.

To conclude this problem description, we summarize the dynamic aspects that characterize the empty container allocation problem: the temporal characteristics of the problem (of various transportation activities, modes and itineraries, port and terminal operations and delays, etc.), the need to reposition (and substitute) containers in order to be able to satisfy future demands, the fact that demands and supplies "appear" at various points in time and vary with time (even for the same "physical" customer), the existence of delivery windows for demand customers, and the stocks of containers at depots used as buffers in the face of varying supplies and demands. Concerning the stochastic aspect of the problem, one notes that it derives from the uncertainties of future supplies and demands. Hence, even though only immediate movements and decisions are of actual interest and are transmitted to the routing module, a much longer planning horizon is required in order to properly represent and assess their impact. Finally, note that the substitution mechanism, the stocks at depots, and the possibility to access containers outside the system (buy, rent, etc.), form a very rich recourse structure for the problem.

The modeling structure that we propose for this problem may be described as a *dynamic network* model over a planning horizon of n days or periods, which is applied in a *rolling horizon* framework. To make for a clearer and smoother presentation, we introduce three formulations. First, a deterministic model of the single commodity case (each container type corresponds to a commodity), followed by the deterministic formulation of the general multicommodity problem where container substitutions are considered. Finally, we initiate the exploration of how to handle uncertainties in the context of container distribution by introducing a stochastic formulation of the single commodity case. Given the current state-of-the-art in the area, as well as the length of the paper, the stochastic formulation of the multicommodity problem clearly is beyond the scope of the present paper (for a possible approach to this problem, see Crainic, Gendreau and Dejax, 1991). All formulations are dynamic, so this characteristic is not included into the section titles.

3. DETERMINISTIC SINGLE COMMODITY MODEL

This section is dedicated to the presentation of the modeling of the problem as a deterministic, dynamic, single commodity formulation. It is an idealized situation which allows nevertheless a good presentation of the notation and mathematical formulation. Most concepts and notation naturally generalize to the other models. Note that if this formulation is to be used *per se*, a number of additional questions should be addressed, including the treatment of the end of the horizon. We discuss this issue in Section 7.

3.1 The Mathematical Model

For a given planning horizon, the configuration of the port and depot network which is to be used by the company is invariant with the specific period (day). On the other hand, the set of "active" customers (i.e. customers with outstanding requests for empties or with available empty containers) does vary according to the time period. Furthermore, requests by the same firm with different due dates have to be treated separately. The set of *identified customer requests* is thus specifically defined for each time period, according to the strategic/tactical plan and the forecast demand for known customer firms. Note that these definitions imply that two requests of the same customer are treated as two different items by the system. Let

T : Length of the planning horizon, $t = 1,2,...,T$.

H : Set of ports (harbors) – entry/exit points of the system.

D : Set of "inland" depots (does not include the port depots).

I^t : Set of *identified demand customer requests*, to be delivered at the latest in period t, $t = 1,2,...,T$.

Δ_i : *Delivery window* of customer i, $i \in I^t$:
if $\Delta_i = 0$, one has to deliver in period t exclusively;
if $\Delta_i > 0$, one has the option of delivering in all periods $t, t-1, ..., t-\Delta_i$.

S^t : Set of *supply customer requests* to be picked up in period t, $t = 1,2,...,T$.

Note that we assign ports and non-port ("inland") depots to different sets; this makes for a cleaner notation, especially for versions of the model presented later in the paper. For plain descriptions however, we shall use the term "depot" to indicate any facility, port or inland depot, where the warehousing function is performed.

The network operations are driven by the amount of empty container supply and demand associated with customers, depots and ports, for each one of the periods of the planning horizon. The supplies and demands of empty containers associated with ports reflect the fact that, in general, any transportation system is embedded into a larger one: regions into states, states into the world, etc. In particular, these values reflect the differences between supplies and demands at the level of the international trade. Let us then define:

X_i^t : Demand of the identified customer i, who requires that containers be delivered at its site in period t, at the latest;
$i \in I^t$, $t = 1,2,...,T$.

X_j^t : Export demand at port j in period t; it represents the empty container requests which originate outside the study area;
$j \in H$, $t = 1,2,...,T$.

Y_s^t : Supply of containers that become available in period t at supply customer s;
$s \in S^t$, $t = 1,2,...,T$.

Y_j^t : Import supply at port j in period t; it represents the input of empty containers from outside the study area;
$j \in H$, $t = 1,2,...,T$.

Generally speaking, the decision variables represent container movements through time and space. Those movements which involve customers are identified by the letter v, while w is used for all others.

In all formality, the variable which corresponds to the allocation decision from depot to customer should be noted

$v_{jt}^{it'}$: Volume of containers allocated from depot j to customer i in period t, and arriving at destination in period t'; $i \in I_j^t$, $j \in D \cup H$, $t = 1,2,...,T$.

to clearly specify the spatial-temporal occurrence that represents the identified customer. However, once the starting time period of a movement is specified, the arriving time is completely determined by the corresponding transit time, and hence it is **not a decision**. The same comments apply to all the decision variables defined therein, and we can thus use a simpler notation.

We use the notation $\tau_{\eta\mu}$ to identify the *transit times* of the various movements, where the origin η and the destination μ may represent a depot, a customer or a port. These delays include the time required for loading at origin and unloading at destination, plus the necessary time to journey between the two points. We assume all transit times to be independent of their starting (t) and arriving (t') periods. We need the following additional notation:

I_j^t : Set of all demand customer requests i which may be satisfied from depot j by a movement which starts in period t and arrives in period $t + \tau_{ji}$; i.e., customer i belongs to I_j^t if $t' - \Delta_i \leq t + \tau_{ji} \leq t'$, for $i \in I^{t'}$;
$j \in D \cup H$, $t = 1,2,...,T$.

$J_i^{t'}$: Set of all depots j which, in period t', may be used to satisfy the demand customer request i, $i \in I^{t'}$; the shipment starts in period t' and arrives in period $t' + \tau_{ji}$ such that $t - \Delta_i \leq t' + \tau_{ji} \leq t$;
$J_i^{t'} \subseteq D \cup H$, $t' = 1,2,...,T$.

J_s^t : Set of all depots j where supply (customer) s, $s \in S^t$, may send in period t the empty containers present on-site; in particular, the port of origin of the loaded movement which generated the current empty supply, noted $j(s)$, belongs to J_s^t;
$J_s^t \subseteq D \cup H$, $s \in S^t$, $t = 1,2,...,T$.

S_j^t : Set of all supply customers s, $s \in S^{t'}$, $t' = t - \tau_{sj}$, from which depot j may receive containers in period t;
$j \in D \cup H$, $t = 1,2,...,T$.

We can now formally define the following decision variables:

v_{ji}^t : Volume of containers allocated in period t, from depot j to customer i. The shipment arrives in period $t' = t + \tau_{ji}$;
$j \in D \cup H$, $i \in I_j^t$, $t = 1,2,...,T$.

v_{sj}^t : Volume of empty containers picked up in period t at supply customer s, and arriving at depot j in period $t + \tau_{sj}$;
$s \in S^t$, $j \in J_s^t$, $t = 1,2,...,T$.

w_{jj}^t : Stock of empty containers available at depot j at the end of period t (the "inventory" flow from depot j in period t to depot j in period $t + 1$);
$j \in D \cup H$, $t = 1,2,...,T$.

w_{jk}^t : Flow of empty containers shipped from depot j to depot k. Shipment starts in period t and is completed in period $t + \tau_{jk}$;
$j \in D \cup H$, $k \in D \cup H$, $t = 1,2,...,T$.

We are now ready to write down the constraints of the deterministic, single commodity model.

Identified Customer Demand.

The demand of an identified customer may be satisfied from all the depots to which the customer is linked, by shipments which arrive in any of the periods that fall in the time interval defined by the customer's delivery window. In some cases, part of the identified demand may already have been satisfied by shipments initiated in periods prior to the beginning of the planning horizon (either because the customer's delivery window starts before this time, or because these shipments have yet to arrive at the customer's site). As these shipments correspond to known decisions, a constant value K_i^t adjusts the identified demand for the planning horizon (of course, the definition of the set $J_i^{t'}$ is modified accordingly):

K_i^t : Volume of containers sent to customer i, $i \in I^t$, before the beginning of the current planning horizon.

The constraint then becomes:

$$\sum_{t' \le t} \sum_{j \in J_i^{t'}} v_{ji}^{t'} = X_i^t - K_i^t \quad \text{for all} \quad i \in I^t,\ t = 1,2,...,T. \tag{1}$$

Note that, to insure feasibility, we assume that for each identified customer demand i, $i \in I^t$, not all the sets $J_i^{t'}$ are empty. In operational terms, this implies that there is always a depot sufficiently close by to satisfy the demand of an identified customer. This corresponds to the fact that the sales department of the company should not accept an order which is temporarily infeasible.

Customer Supply.

Empty containers, which are available at a customer site, have to be moved out, and may be sent either to the port where the initial loaded movement originated, or to one (or several) of the depots to which the customer is assigned in the strategic/tactical plan:

$$\sum_{j \in J_s^t} v_{sj}^t = Y_s^t \quad \text{for all} \quad s \in S^t, \quad t = 1,2,...,T. \tag{2}$$

Stocks at Non-Port Depots.

For each depot, the stock of empty containers is computed at the end of each time period. Out of the depot are the volumes shipped to the other depots, to ports and to customers. The volume input during the period is made up of container traffic that left in previous periods from the other depots, the ports and the supply customers, and that is arriving to the depot in the current period, according to the corresponding transportation times. This volume is added up to the stock available at the beginning of the period, and to the containers which are brought in from outside the system:

b_j^t : Volume of containers "borrowed" (from partner companies; it may also represent rented containers, or newly bought ones) at depot j in period t;
$j \in D \cup H$, $t = 1,2,...,T$.

We denote by w_{jj}^0 the stock available at depot j, $j \in D \cup H$, at the beginning of the planning horizon. As previously, we use a constant to adjust the depot supply for the containers shipped to the depot before the beginning of the current planning horizon:

K_j^t : Volume of containers sent to depot j before the beginning of the current planning horizon and arriving in period t;
$j \in D \cup H$, $t = 1,2,...,T$.

Then

$$w_{jj}^t = w_{jj}^{(t-1)} + \sum_{t' \le t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{t'} + \sum_{s \in S_j} v_{sj}^{t'} \right] + b_j^t + K_j^t \tag{3}$$

$$- \sum_{k \in D \cup H} w_{jk}^t - \sum_{i \in I_j} v_{ji}^t \quad \text{for all} \quad j \in D, \quad t = 1,2,...,T .$$

Port Supply, Demand and Stock.

$$w_{jj}^t = w_{jj}^{(t-1)} + \sum_{t' \le t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{t'} + \sum_{s \in S_j^{t'}} v_{sj}^{t'} \right] + b_j^t + K_j^t \qquad (4)$$

$$- \sum_{k \in D \cup H} w_{jk}^t - \sum_{i \in I_j^t} v_{ji}^t - X_j^t + Y_j^t + e_j^t \quad \text{for all } j \in H,\ t = 1,2,...,T\ .$$

where

e_j^t : Unsatisfied export demand of empty containers in port j, in period t ($e_j^t \ge 0$); $j \in H$, $t = 1,2,...,T$.

The constraints determining the stock levels at port depots are very similar to those controlling stocks at non-port depots, with the exception that the net import/export demand at the port has to be met. We identify the slack variable as the *unsatisfied demand*, mainly to account for the cost of not meeting, in some cases, the export demand. This is the only demand we assume possible to reject. Note that the presence of both the b_j^t and e_j^t variables, with the corresponding costs, allows one to perform a number of interesting analyzes on the trade-offs between the costs of not meeting the export demand versus the cost of bringing in new containers into the system.

Balancing Flows.

$$l_{jk}^t \le w_{jk}^t \le u_{jk}^t \quad \text{for all } j \in D \cup H,\ k \in D \cup H,\ t = 1,2,...,T\ . \qquad (5)$$

where

l_{jk}^t (u_{jk}^t) : lower (upper) bound on a balancing movement of containers initiated in period t, from depot j to depot k;
$j,k \in D \cup H$, $t = 1,2,...,T$.

These equations insure that balancing movements are carried out according to the rules of the company, especially its agreements with carriers. In particular, the l_{jk}^t and u_{jk}^t bounds are exogeneous to the model and reflect these policies and agreements.

To write the objective function of the model, we need to define the "cost" criteria associated with each one of the transportation and warehousing activities defined in the system.

In all generality, the cost function associated with a specific activity (e.g. moving containers, in a specific time period, between a given customer and a particular port) could depend upon the values of all flow variables in the model. Yet, in practice, such a broad definition is not required, since most empty traffic that involves customers takes place over relatively short distances and moves only a few containers at one time (this type of movement is almost always truck based). We then assume that for this type of activity, the cost function depends only on the intensity of the activity. Furthermore, in order to keep the presentation simple, we assume that all functions are linear. We define the following notation:

c_{ji}^t : Unit transportation cost from depot j to demand customer i, in period t;
$j \in D \cup H$, $i \in I^{t'}$, $t = 1,2,...,T$.

c_{sj}^t : Unit transportation cost from supply customer s to depot j in period t;
$s \in S^t$, $j \in D \cup H$, $t = 1,2,...,T$.

c_{jk}^t : Unit transportation cost from depot j to depot k in period t;
$j \in D \cup H$, $k \in D \cup H$, $t = 1,2,...,T$.

c_j^t : Cost of holding one container at depot j in period t;
$j \in D \cup H$, $t = 1,2,...,T$.

$\bar{c}_j^t$: The cost of bringing in one container from outside the system, at depot j, in period t;
$j \in D \cup H$, $t = 1,2,...,T$.

$\tilde{c}_j^t$: Unit penalty cost for not satisfying the external demand of empty containers at port j in period t;
$j \in H$, $t = 1,2,...,T$.

The total cost of operating the land distribution and transportation system may then be written as:

$$\sum_{t=1,2,...,T} \left\{ \sum_{j \in D \cup H} \left[\sum_{i \in I_j^t} c_{ji}^t \, v_{ji}^t + \sum_{k \in D \cup H} c_{jk}^t \, w_{jk}^t + c_j^t \, w_{jj}^t + \bar{c}_j^t \, b_j^t \right] \right. \tag{6}$$

$$\left. + \sum_{j \in H} \tilde{c}_j^t \, e_j^t + \sum_{s \in S^t} \sum_{j \in J_s^t} c_{sj}^t \, v_{sj}^t \right\}$$

and the model minimizes (6), subject to constraints (1) to (5) and the usual non negativity restrictions on the decision variables.

3.2 The Network Representation

The problem we study displays a natural network structure, which is present in the deterministic model of this section, as well as in the more complex formulations that follow. As a basis for these formulations, we introduce the minimum cost network flow model that corresponds to the single product deterministic case. The transformation involves the construction of the network and the specification, for each arc, of the flow variables, of the corresponding flow upper and lower bounds and of the cost function. The usual flow conservation constraints may be then directly inferred. However, when the network variables and constraints correspond rather straightforwardly to those of the previous model, we do not explicitly rewrite them.

The formulation is based on a multisource multisink network $G = (N, A)$, whose spatial topology conceptually unfolds in two dimensional space: *time* spans the x axis, while *space (location)* unrolls on the y axis. The time axis is divided into periods, and for each period a network representation of the actual distribution and transportation system of the company unfolds on the space axis. Although the sets of ports and depots used by the company do not vary within the planning horizon, this representation is specific to each particular period, since the sets of identified supply and demand customer requests are usually different for each time occurrence. "Dynamic" arcs link geographical representations along the time axis, and these arcs may span more than one time period, as will be shown below.

The node set N encompasses elements which represent the components of the distribution system: ports, depots and customers. The sets of nodes representing identified demand and supply customer requests are specific to each time period; hence, we use the I^t and S^t notation previously defined. The ports and depots used by the company are the same for each specific period; hence a replication of these elements (noted H^t and D^t, respectively) is required for each time occurrence of the spatial representation. Note that in this representation, a large number of nodes appear as sources or sinks to account for the containers that enter and exit the system; a limited number of the X, Y and K constants associated with such events are indicated in Figure 3.

Individual nodes are identified by the juxtaposition of the letter identifying the network element (depot, port, demand or supply customer request) and the time period. Thus, in the

example shown in Figure 3, $k(t-2)$ and $i(t)$ correspond to the replication of depot k in time period $t-2$, and to demand customer request i, $i \in I^t$, respectively.

Finally, we add two super-source nodes to the set N, to account for the previously defined relationships with the external world:

- α : Source for the "new" (borrowed from partner companies, rented, bought, etc.) containers introduced into the system:
- α' : Dummy node defining arcs that capture the unsatisfied external demand at ports.

The definition and inclusion of the links of the network is mainly performed according to the company's operating policies, the strategic/tactical plan, the ship schedules and the temporal characteristics of supplies and demands. The plan specifies to which depot(s) each customer may be linked, and the permitted depot to depot movements. The ship schedules directly determine when exchanges with the outside world occur and influence, through the forecasting process, when supplies and demands appear in the system. Additional arcs are included to model warehousing decisions, and to account for specific preoccupations of day-to-day operations.

We illustrate the network representation by using the simple case of Figure 2. The resulting network is displayed in Figure 3 for three time periods: one harbor and two inland depots, each replicated three times, plus two identified demand customer request nodes, $i(t)$ and $\hat{i}(t-1)$, and one identified supply customer request node $s(t-2)$. It should be noted, however, that in order to keep the presentation clear, we have drawn only "representative" arcs, instead of the complete network. (When no risk of ambiguity exists, in the following we only define the notation for period t.)

The arcs $(h(t), k(t))$ and $(k(t), h(t))$ are shown in order to illustrate the possibility of limited exchanges between "neighboring" depots. These exchanges require small transportation times and involve relatively limited volumes of containers. Such movements may be initiated at any moment, for example, when the transportation cost from depot h to depot k plus the distribution cost starting from k is lower than the holding cost at h plus the corresponding distribution cost. Although the flow variables are similarly identified, this type of movement

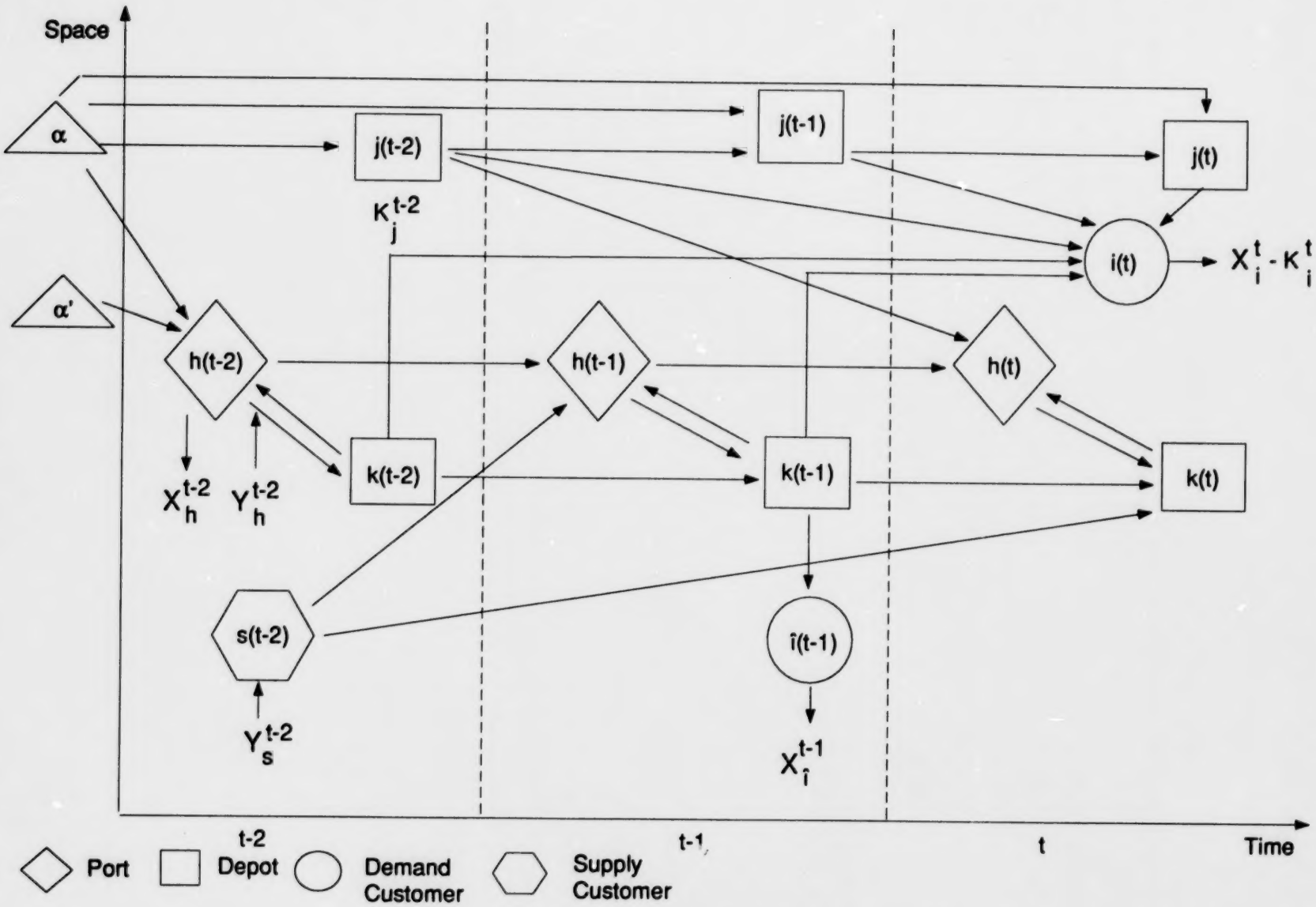

Figure 3. Outline of the Time-Space Network Representation

is different from the massive interdepot balancing traffic, represented by arcs spanning more than one period, such as arc $(j(t-2), h(t))$ of Figure 3.

The depot to customer arcs, of type $(j(t'), i(t))$, illustrate the fact that a customer demand may be met from several, close or not so close, depots, and that the requested containers may arrive at any moment during the specified delivery window. Thus, for example, the demand of customer i, which may be served from depots j and k, has to be delivered in period t with a delivery window $\Delta_i = 2$. Depot j may deliver in less than a period; hence the three possible arcs, $(j(t-2), i(t))$, $(j(t-1), i(t))$ and $(j(t), i(t))$, are included into the network. Depot k is farther away and thus cannot make a timely delivery by allocating in period t; hence, link $(k(t), i(t))$ has not been included. On the other hand, node $\hat{i}(t-1)$ represents the case of a customer with no delivery window $(\Delta_{\hat{i}} = 0)$ and only one depot that is sufficiently close to be able to satisfy the request (thus implying $K_{\hat{i}}^{t-1} = 0$).. We thus include into the dynamic network a $(j(t'), i(t))$ depot to customer arc only if the corresponding transportation operation permits to meet the demand in time, i.e. if $t - \Delta_i \leq t' + \tau_{ji} \leq t$. Note that, since the arc originates in the time period when the transportation operation is initiated from the depot and ends in the latest delivery time period for the customer, the number of periods the link spans may be greater than the actual transportation time.

We have also illustrated in Figure 3 some of the "holding" arcs, of type $(j(t), j(t+1))$, which model the management of inventories of empty containers at depots, as well as the two super-sources, α and α'. This enables us to illustrate the arcs used to model the various relationships with the outside world. Arc $(\alpha', h(t-2))$ supports the flow, represented by the slack variable $e_j^{(t-2)}$, required to meet the external demand for empty containers that the system cannot directly satisfy. Such arcs are introduced into the network only in periods when external demands are specified at port nodes. Arcs of the type $(\alpha, h(t))$ and $(\alpha, j(t))$ (only four such arcs are drawn in Figure 3) capture the recourse to external sources of empty containers in order to satisfy the total demand. Although in theory the upper flow limit on these arcs should be equal to infinity (in order to insure feasibility), in practice some sensible and practical finite value bounds exist. Suitable total supplies have to be defined at nodes α and α'.

The arcs $(s(t-2), h(t-1))$ and $(s(t-2), k(t))$ implement the requirement that empty containers have to be picked up at a (identified supply) customer's location in order to be

transported either to the port of origin of the initial loaded movement, or to one (or several) of the designated depots. Note that arcs between customers and depots span as many periods as required by their respective transportation times. Also note, that one may quite straightforwardly model the possibility that containers may be left at customers' sites for later removal: one simply replicates the corresponding $s(t)$ nodes the number of periods the containers may be left there, and adds *customer holding* arcs $(s(t), s(t+1))$ capturing the effect of keeping inventories at customers' sites. Of course, appropriate cost functions have to be defined since, otherwise, significant volumes of containers might be left at customers' locations in order to avoid inventory costs at depots. This approach assumes that one does not need to pick up all containers present at a given location at the same time.

Finally, Figure 3 illustrates the demands and supplies associated with the nodes of the network. It may be noted that port supplies and demands appear only at those periods when ships are scheduled. Otherwise, the supplies and demands of the network model are the same, and appear at the same periods, as those of the previous formulation. In particular, supplies and demands are adjusted to account for the activities of previous periods, when necessary.

4. DETERMINISTIC MULTICOMMODITY FORMULATION

The possibility to substitute one type of container for another increases the flexibility of the distribution system and enhances its ability to adequately respond to the customers' requests. Hence, a more general model of the container distribution problem is a multicommodity formulation, where each container type is represented by a different commodity, and where partial substitutions are allowed among the various container types.

Let P be the set of container types (products), with $p \in P$. Then, in general, the notation of the preceding section generalizes rather straightforwardly to the multicommodity case by inserting the product index p before the time index t. Hence, for example, the identified customer demand becomes:

X_i^{pt} : Demand of identified customer i, who requires that containers of type p be delivered at its site in period t at the latest;
$i \in I^{pt}$, $p \in P$, $t = 1,2,...,T$,

while the corresponding node set is written I^{pt}. For the sake of brevity, we do not explicitly redefine all variables, constants and sets, nor do we write down all the equations previously described. We indicate hereafter only those elements that change or that are new due to the multiproduct nature of the problem.

The network $G = (N, A)$ that underlies the deterministic formulation is now a three dimensional structure: *time* on the x axis, *space (location)* on the y axis, and *products (container types)* on the z axis. The product axis has as many units as the number of container types defined in the problem. The network representation of the actual distribution and transportation system of the company, which unfolds on the space axis, is now specific to each particular *(time, product)* occurrence, since the sets of supply and identified demand customer requests usually vary with the container type as well as with the time period. Note that the sets of ports and depots used by the company continue to represent the invariant part of the network. "Dynamic" arcs link geographical representations along the time axis, while substitution arcs link the various commodity networks for a given time period.

Substitution factors are used to represent some of the technical restrictions related to using a container of a type different from the one required by the customer. For any given

pair (p, r) of products, the substitution factor a_{pr} represents the fact that more than one container of type p might be required to replace one container of type r. It should be stressed that these factors do not correspond to direct volume relationships between container types; they rather are the inverse of some integer number:

$$a_{pr} = (\text{ number of type } p \text{ containers needed to replace one type } r \text{ container })^{-1}.$$

This approach is general although its implementation may require a detailed definition of container types, and the specification of demands in appropriate units. For instance, one may have to differentiate between demands that have to be satisfied by 40' containers and those which may be accommodated with either 40' or 20' ones.

To adequately represent the substitution mechanism and its consequences, we distinguish the containers which "physically" are of a given type p, from the *equivalent containers of type p* which are to be distributed to customers. This is tantamount to differentiating the two main depot functions: managing stocks of containers of various types and allocating containers to satisfy customer demands. As a modeling principle, substitutions occur only at depots. Furthermore, stocks of containers at depots, all interdepot traffic, as well as all exchanges with the outside world, are of true containers of the given type, while demands are satisfied by using equivalent containers.

In the network formulation, this approach is implemented by associating two nodes to each depot and port:

D^{pt} (H^{pt}) : Subset of nodes representing the replication, for container type p and period t, of the physical locations (depots and ports) where stocks of empty containers are managed; this corresponds to the *warehousing* function of the depot; $p \in P$, $t = 1,2,...,T$.

$\tilde{D}^{pt}$ $(\tilde{H}^{pt})$: Subset of nodes representing the replication, for product p and period t, of the source depots (ports), from where equivalent empty containers of type p are allocated to customers; this corresponds to the *allocation* function of the depot; $p \in P$, $t = 1,2,...,T$.

This definition of the network structure ensures that containers that are substituted are directly used to satisfy demands and do not end up in inventories at depots. The customer

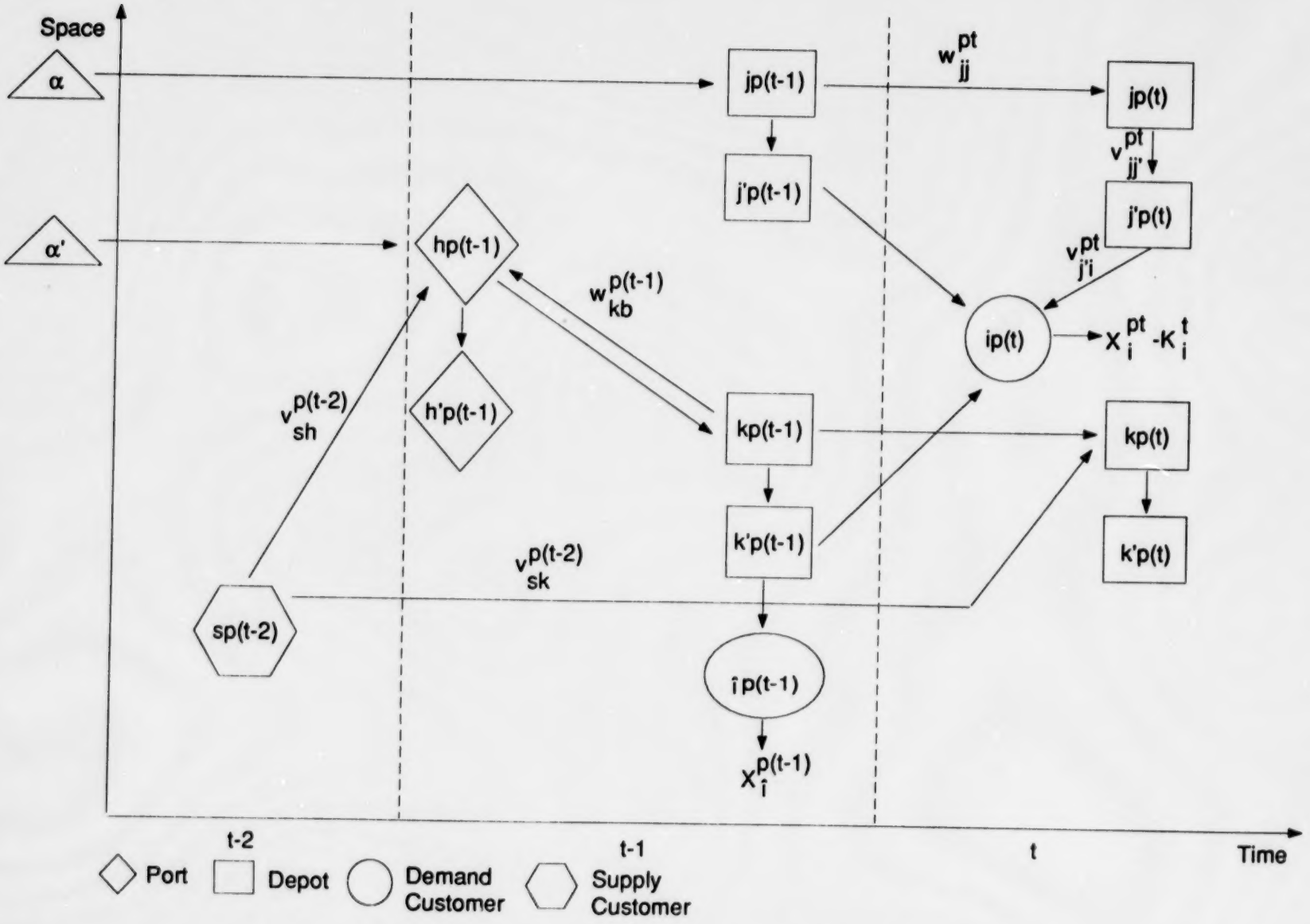

Figure 4. Outline of a Time-Space Container Type Layer

allocation decision variable then becomes

v_{ji}^{pt} : Volume of equivalent containers of type p allocated in period t from depot j to customer i; $j \in D \cup H$, $i \in I_j^{pt}$, $p \in P$, $t = 1,2,...,T$;

while new decision variables have to be defined:

$v_{jj'}^{pt}$: Total volume of containers of type p, in stock at depot j, which will be allocated, in period t, as equivalent containers of type p;
$j \in D \cup H$, $p \in P$, $t = 1,2,...,T$;

$w_{jj'}^{prt}$: Number of empty containers of type p substituted for containers of type r, at depot j in period t;
$j \in D \cup H$, $p \in P$, $r \in P$, $t = 1,2,...,T$.

Figures 4 and 5 illustrate these concepts. For clarity reasons, Figure 4 only displays part of the subnetwork that corresponds to the space-time layer of product p. It also shows a few of the corresponding decision variables. The example network of Section 3 is also used here. Nodes are noted by the juxtaposition of the letter that identifies the network component, the product and the time period.

As illustrated in Figure 4, for each container type layer in the network and for each period, all flows between "warehousing" nodes (depots and ports in H^{pt} and D^{pt}, respectively) represent movements of true containers. The flow on arc $((hp(t-1), (kp(t-1))$ is a case in point. True containers also flow on the arcs from each "warehousing" depot to its corresponding "allocation" one, such as arc $((jp(t), j'p(t))$. On the other hand, flows that originate from the "allocation" depot nodes, and that end up at demand customer request nodes, e.g. on arc $(j'p(t), ip(t))$, should be interpreted as movements of equivalent containers.

Movements originating from supply customer request nodes (e.g., on arcs $(sp(t-2), kp(t))$ and $(sp(t-2), hp(t-1))$), warrant a special note. Indeed, since substitute containers may have been used for the original loaded movement, technically, these flows could be interpreted as representing traffic of "equivalent" containers. This approach would rapidly put the system into a chaotic state, as "equivalent" and "physical" containers would mix into the depot inventories. An elegant way of addressing this issue is based on the

observation that, since supply customer requests are generated from actual movement orders for loaded containers, one knows exactly the container types involved when these requests materialize. Hence, at generation time, the supplies at a customer's site may be transformed into as many different supply customer requests as the number of different container types present at the customer's site. Consequently, the supply customer to depot flows may be assumed to represent traffic of true, "physical", containers.

Permitted substitution activities are represented by arcs linking different container type layers of the network. At a depot j, substitution arcs $(jp(t), j'r(t))$ start at the warehousing node of one product (e.g., p) and point to the allocation node of another (e.g., r). Flow variables $w_{jj'}^{prt}$ measure the volume of substitutions, which are governed by the substitution (gain) factors a_{pr}. Upper bounds B_j^{prt} limit the amount of flow transferring from one commodity layer to another, thus modeling the actual restrictions present in company operating policies.

Figure 5 shows a number of such arcs which illustrate some of the possible substitution cases. Note that the $jp(t)$ and $j'p(t)$ nodes of Figure 5 are the same as those of Figure 4, while the three other node pairs represent the warehousing and allocation activities in period t for three other products (l, r and g) at the same depot j. Figure 5 also displays a number of arc decision variables and substitution factors.

Three cases are illustrated. On one hand, containers of type p and r may be substituted one for the other; hence both arcs $(jp(t), j'r(t))$ and $(jr(t), j'p(t))$ exist. On the other hand, containers of type p and r may be substituted for the type l, but the converse operation is not allowed. Finally, no substitutions are permitted between containers of type g and any of the types p, r or l, and no substitution arcs are defined.

Relatively few modifications have to be performed to the constraints of the single commodity model, besides the insertion of the product index p. Indeed, only constraints (3) and (4) have to be modified to account for the containers that are used for substitutions

$$w_{jj}^{pt} = w_{jj}^{p(t-1)} + \sum_{t' \le t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{pt'} + \sum_{s \in S_j^{pt'}} v_{sj}^{pt'} \right] \qquad (7)$$

$$+ b_j^{pt} + K_j^{pt} - \sum_{k \in D \cup H} w_{jk}^{pt} - v_{jj'}^{pt} - \sum_{r \in P} w_{jj'}^{prt}$$

$$\text{for all } j \in D,\ p \in P,\ t = 1,2,\ldots,T,$$

and

$$w_{jj}^{pt} = w_{jj}^{p(t-1)} + \sum_{t' \le t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{pt'} + \sum_{s \in S_j^m} v_{sj}^{pt'} \right] \tag{8}$$

$$+ b_j^{pt} + K_j^{pt} - \sum_{k \in D \cup H} w_{jk}^{pt} - v_{jj'}^{pt} - \sum_{r \in P} w_{jj'}^{prt} - X_j^{pt} + Y_j^{pt} + e_j^{pt}$$

$$\text{for all } j \in H,\ p \in P,\ t = 1,2,...,T,$$

respectively, while a new set of constraints has to be defined to represent the

Depot Allocation Balance:

$$v_{jj'}^{pt} + \sum_{r \in P} a_{rp}\, w_{jj'}^{rpt} - \sum_{i \in I_j^m} v_{j'i}^{pt} = 0 \qquad \text{for all } j \in D \cup H,\ p \in P,\ t = 1,2,...,T. \tag{9}$$

Most of the cost definitions of the previous sections readily generalize to the multiproduct case, especially since we assume linear cost functions. Define

c_j^{prt} : Substitution cost, at depot j in period t, of a container of type r by a container of type p;
$j \in D \cup H,\ p \in P,\ r \in P,\ t = 1,2,...,T.$

The deterministic multiproduct model then minimizes

$$\sum_{t=1,2,...,T} \sum_{p \in P} \left[\sum_{j \in D \cup H} \left(\sum_{i \in I_j^m} c_{ji}^{pt}\, v_{j'i}^{pt} + \sum_{k \in D \cup H} c_{jk}^{pt}\, w_{jk}^{pt} \right.\right. \tag{10}$$

$$\left.\left. + c_j^{pt}\, w_{jj}^{pt} + \bar{c}_j^{pt}\, b_j^{pt} + \sum_{r \in P} c_j^{prt}\, w_{jj'}^{prt} \right) \right.$$

$$\left. + \sum_{j \in H} \bar{c}_j^{pt}\, e_j^{pt} + \sum_{s \in S^m} \sum_{j \in J_s^m} c_{sj}^{pt}\, v_{sj}^{pt} \right]$$

subject to constraints (1), (2), (7), (8), (9) and (5), to upper bounds on the substitution variables, and the usual non negativity restrictions on the decision variables.

To solve this large dynamic linear multicommodity minimum cost generalized network flow problem, one could apply a standard network flow with gains algorithm to the multilayer

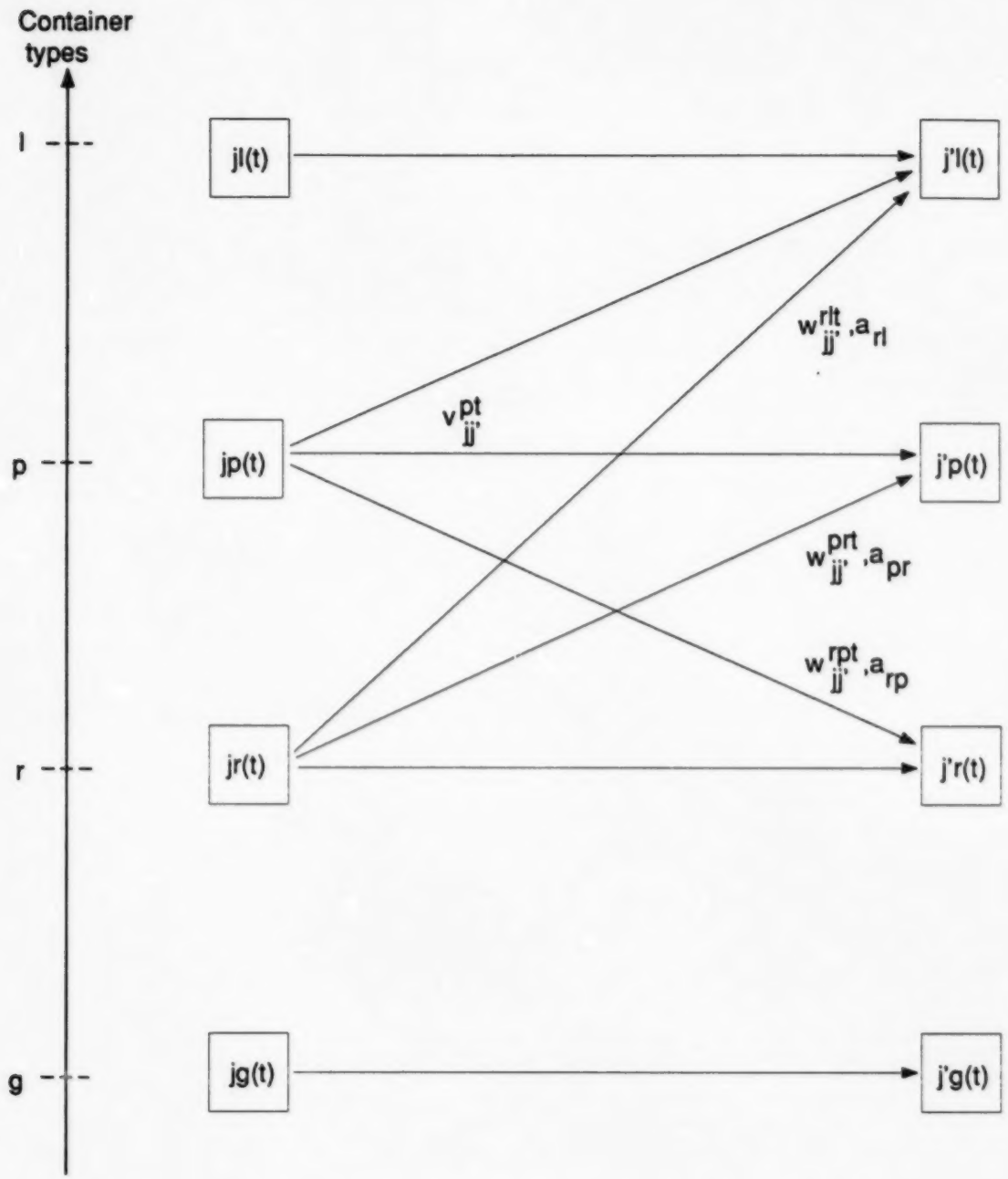

Figure 5. Partial Substitution Subnetwork :
Four Commodity Layers at Depot j in Period t

representation of the network, thus solving the problem as a single commodity one. However, this approach does not account for the very special structure of the network. More precisely, it does not acknowledge the fact that, were it not for the substitution arcs, the problem would decompose naturally in a series of much smaller pure network problems. In this context, decomposition methods which seek to exploit this near-separability of the problem seem quite attractive, and we intend to follow this approach.

5. STOCHASTIC SINGLE COMMODITY MODEL

This section is dedicated to the development of a formulation of the dynamic single commodity case that takes into account the stochastic nature of some of its major components. We aim at a formulation that is both analytically tractable and numerically solvable, and that parallels as closely as possible the deterministic model of Section 3.

In previous sections, we assumed all demands and supplies to be perfectly known throughout the planning horizon. We now relax this assumption and distinguish between two types of demands and supplies: on the one hand, deterministic demands and supplies resulting from firm requests made prior to the beginning of the planning horizon and, on the other hand, stochastic demands and supplies corresponding to forecast but uncertain future requests. In the dynamic context of a planning horizon of T periods, it is also assumed that the first period of the horizon, for which decisions have to be implemented immediately, is fully deterministic. The uncertainties thus appear only in periods 2 to T and, in a rolling horizon mode of operation, their main role is to serve to ponder today's decisions by their estimated effect on future ones. It is also important to note that, in this section, the durations of the various transportation activities are still assumed to be deterministic parameters.

The remainder of the section is organized as follows. We first discuss in more detail the stochastic demands and supplies. We then briefly review possible modeling approaches to the problem, before presenting a tractable two-stage restricted recourse model for it. This is followed by a complete mathematical formulation of the model, and by a discussion of algorithmic issues relative to its solution.

5.1 Stochastic Demands and Supplies

The stochastic elements we consider are associated with the "internal" customer demands and supplies, as well as with the demands and supplies of empty containers from outside the system which occur at exit and entry points, ports in particular.

The internal stochastic demand is made up of uncertain requests associated with given customers. As in the previous sections, each of these requests should specify the customer location, the number of containers demanded and the delivery window. It should be noted that, aside from "real" customers of the company, other events which have an impact on the

stock of containers, such as accidents temporarily or permanently damaging containers or requests from partner companies, can easily be modeled as demand customer requests within this framework.

In a similar fashion, the internal stochastic supply is made up of uncertain requests for picking up available empty containers at customer sites in future periods. Events such as returns of repaired or lent containers may also be modeled as supply customer requests. It is important to note, however, that for supply customer requests the uncertainty derives, in most cases, not from the randomness of the amount of containers available for pickup, but rather from the "when does that supply become available" type of question. Indeed, most supply customer requests are generated based on the knowledge, gained from precise import port to customer movement orders, that a given number of loaded containers are present, or are arriving, at a customer's location. Hence, while the volume of supply available at a customer's location is quite well known, the moment when the empty containers will be available *becomes a random event*. Indeed, if loaded containers are delivered today, this does not necessarily mean that the containers will be emptied by, say, the following day. It may be tomorrow, or the day after, or the day after that. In fact, one may infer, from historical data, the amount of variation particular to a given customer, variation that may increase when the loaded containers have not yet arrived at their destination.

Finally, the external supplies and demands of empty containers associated with ports in future periods may be also viewed as random variables which reflect, in particular, the differences between supplies and demands at the level of the international trade.

5.2 Possible Modeling Approaches

The deterministic model of Section 3 describes fairly extensively the complex relationships existing between the various decisions that have to be made when addressing the problem of optimally allocating containers over a large network. Ideally, we would like to be able to come up with a stochastic model which would account for the uncertain nature of supplies and demands, while retaining as much as possible of the interactions between the decision variables. At the same time, we would also like this model to be computationally tractable for large size networks. As we shall see further on, these objectives are unfortunately contradictory. The purpose of this section is to examine a series of possible formulations in which

different trade-offs are made between the complexity of the interactions and the tractability of the models.

We first consider a direct adaptation of the model of Section 3. Without going into all the intricacies of defining exactly when the random variables describing stochastic demands and supplies are realized, let us simply suppose that new information on these becomes available at the beginning of each period. Under such an assumption, it should be possible to revise the allocation of containers at the beginning of each period, and the natural way to state the problem is as a *T-stage stochastic program with full network recourse* (this indicates that all future decisions are re-optimized in each period). Given the complex relationships linking the decision variables in different periods, the simplest way to write down this model is by resorting to a *scenario formulation*, in which each scenario corresponds to a specific joint realization of the random variables. Such a formulation yields a large mathematical program in which all decision variables and constraints have to be replicated for each scenario, while the replications of the decision variables are linked by a set of *non-anticipativity constraints* which insure that the decisions made in any stage t depend only on the values of the random variables realized up to that time (see Wets, 1974, and Rockafellar and Wets, 1991, for further details). Unfortunately, the size of most container land distribution networks is such that the number of possible joint realizations of the random demands and supplies is extremely large. This implies that, even if only a small fraction of these possible realizations were sampled as scenarios, the resulting mathematical program would still be numerically intractable.

One way to somewhat simplify the previous model would be to assume that all the information regarding uncertain supplies and demands becomes available at the beginning of the second period (all random variables would thus be realized at that time). This yields a *two-stage stochastic program with full network recourse* in which periods 2 to T are lumped together in the second stage. Such a simplification however does not significantly reduce the computational burden since each evaluation of the recourse function associated with the second stage, for a given set of values of the first period decision variables, still involves solving a deterministic problem with $(T - 1)$ periods for each scenario considered.

If we want to achieve a tractable model, we must therefore accept much stricter restrictions regarding the recourse strategy. Hence, the approach that we have decided to use in this paper is the following: all the "real" decisions are made at the beginning of the planning

horizon on the basis of the information available at that time, and we let the system under study simply to react in a pre-specified fashion to "adjust" these decisions to the random events occurring in later periods as these unfold (decisions in these periods are not re-optimized). More formally, the original problem is divided into two *stages*. In the first stage, we consider a deterministic optimization problem encompassing the decisions made at the beginning of the planning horizon, in particular, the first-period decisions. In the second stage, we determine, for given values of the first stage variables and of the random parameters, the impact on the overall objective of the corrective actions undertaken in reaction to random events. Such an approach obviously limits the range of actions in future periods, but it allows us to account for uncertainties with a reasonable computational effort. It should also be pointed out that, if the model is properly defined, it should provide good conservative estimates of the costs to be incurred in the later periods (one could certainly do better on the average by re-optimizing decisions on the basis of the new information which becomes available as time goes on).

5.3 The Two-Stage Restricted Recourse Model

To fully specify the model, we must spell out how the random parameters will be defined, the set of first stage decision variables, and how the container allocation "system" will handle random events. Let us first examine in this context, the stochastic elements previously described: uncertain customer and export supplies and demands.

With regards to uncertain customer requests, we face a difficult problem due to the fact that these may, in some cases, be satisfied from more than one depot. When such a request is realized, we should thus have to decide by which depots it will be handled. In the case of demands, this is further compounded by the necessity to make a decision on the timing of the container movements to the customer location (because of the flexibility afforded by the delivery windows). Such a latitude of action is hardly compatible with the requirements that we have to impose on the recourse strategy. To untangle the situation, we first force the assignment of each customer to a single depot for its uncertain requests of a given type (i.e., supply or demand) over the planning horizon. This is not, in general, such a big distortion, since most customers will already be assigned to a single depot for a given type of containers and requests as a result of planning at the strategic/tactical level. As for delivery windows, we

can adopt, in a stochastic setting, the conservative policy of shipping containers to customers as late as possible, i.e., at the end of the forecast delivery windows. Having already forced the assignment of customers to a specified depot for uncertain demands, the period in which a shipment should occur under this policy can easily be determined.

Alternate hypotheses may be adopted, corresponding to different policies, as long as an uncertain supply or demand is associated to only one depot and one time period. However, once such an hypothesis is in effect, uncertain demands and supplies for each depot and time period can be aggregated into a single random variable for each type of movement:

χ_j^t : Random demand in the region of depot j, in period t;
$j \in D \cup H$, $t = 2,...,T$.

ψ_j^t : Random supply of containers available in period t at depot j;
$j \in D \cup H$, $t = 2,...,T$.

Deriving the probability distributions of these variables requires careful analysis, but we defer the discussion relative to this issue to Section 7, and simply assume for the time being that these distributions are known.

Apart from customer requests, containers are received at ports from outside the system and export demands for empty containers are also made at these locations. To simplify the exposition, we assume that, except for the first period in which such supplies and demands should be known with certainty (we retain the notations X_j^1 and Y_j^1, for $j \in H$, for these), these are random variables, denoted

ξ_j^t : Export demand at port j in period t; $j \in H$, $t = 2,...,T$.

ϕ_j^t : Import supply at port j in period t; $j \in H$, $t = 2,...,T$.

Having spelled out the random parameters describing uncertain events in our representation of an empty container distribution system, we now turn to the definition of the first-stage problem. First, as already mentioned, all decisions in the first period of the horizon should be included in this stage. As for later periods, it should be noted that certain customer requests are not restricted to the first period. Indeed, in real-life applications, one would expect a fair portion of firm customer requests to be known some periods in advance. For these, it seems natural to come up with firm allocation decisions right at the outset of the planning horizon,

and thus to include the decisions concerning the movements used to satisfy them in the first-stage problem. Second, it was seen in the deterministic model of Section 3 that depot-to-depot movements are largely ruled by exogeneous constraints. In such a context, it seems preferable to make all such movements deterministic, and thus to include them into the first-stage problem.

To summarize, the first-stage problem covers all first period decisions, as well as all depot-to-depot flows and movements relative to identified certain customer requests in later periods. Given these, the second-stage model is simply a container inventory model that tracks down hoe the stocks of available containers at depots in periods 2 to T will fluctuate in reaction to random supplies and demands. We now specify exactly how this process will take place. (Note that, in the following, all Greek letters refer to random variables.)

Consider a given depot j at the beginning of a period t ($t \geq 2$). In period t several events may occur that will impact the level of the stock of containers on hand:

1) There is a net inflow or outflow of empty containers arising from the deterministic decisions of the first-stage problem; let $\tilde{w}_j^t$ denote this inflow (outflow, if negative).
2) A random supply (ψ_j^t) is received from uncertain supply customers.
3) There is a random demand (χ_j^t) from uncertain demand customers.
4) If depot j is a port, a random import supply (ϕ_j^t) is received, and there is a random export demand (ξ_j^t) for empty containers.

In the deterministic models of the previous sections, it was always assumed that customer demand had to be met, but that export demand could be left unsatisfied if it was more economical to do so. Such a latitude is not easily accommodated in our restricted recourse framework. Hence, in the present stochastic model, we make the simplifying assumption that all demands have to be satisfied exactly. This allows us now to aggregate all random demands and supplies into a single random variable for each depot and time period:

Λ_j^t : Net random demand (supply, if negative) at depot j, in period t;
$j \in D \cup H$, $t = 2,...,T$,

and we have

$$\Lambda_j^t = \begin{cases} \chi_j^t - \psi_j^t, & j \in D \\ \chi_j^t - \psi_j^t + \xi_j^t - \phi_j^t, & j \in H. \end{cases}$$

It should be noted that we were able to perform this aggregation because of the simplifying assumptions made earlier, and because we are dealing with a single type of containers; this would not be possible if we had different rules for handling export demands or several container types with possible substitutions.

Given these elements, there are only two things that one may do at a depot: "borrow" containers from external sources, and hold on to some of the containers for later periods (these two actions correspond respectively to the b_j^t and w_{jj}^t decision variables of the deterministic formulation). To represent these activities, we introduce the following second-stage variables:

ω_j^t : Stock of empty containers available at depot j at the end of period t; $j \in D \cup H$, $t = 2,...,T$.

β_j^t : Volume of containers borrowed from external sources at depot j in period t; $j \in D \cup H$, $t = 2,...,T$.

We can then write the stock balance equation

$$\omega_j^t = \omega_j^{(t-1)} + \tilde{w}_j^t + \beta_j^t - \Lambda_j^t.$$

Note that, since the values of the second-stage variables for any depot j in any period $t \geq 2$ depend solely on the inventory levels in previous periods, and thus on w_{jj}^1, the deterministic inventory at the end of period 1, as well as on the net inflows ($\tilde{w}_j^t$) and net random demands (Λ_j^t) for periods $t = 2,...,T$, the second-stage is **separable by depot**, and the overall recourse function of the problem can be written as $\sum_{j \in D \cup H} Q_j(w_{jj}^1, \underline{\tilde{w}}_j)$, where $Q_j(w_{jj}^1, \underline{\tilde{w}}_j)$ is the recourse function for depot j, while $\underline{\tilde{w}}_j$ represents the vector of $\tilde{w}_j^t$ variables for periods $t = 2,...,T$.

Before closing this section and proceeding to the formal description of the stochastic model, two remarks are in order. First, it should be noticed that we have lost the exact location of customer sites by performing the aggregation of uncertain demands and supplies. It is

thus no longer possible to compute the exact cost of movements between these sites and their assigned depots. But since these movements are "forced" in the model, they have no impact on the optimization; they can therefore be dropped from the objective function. Second, readers familiar with stochastic programming will have noticed that the recourse structure that we have defined is not the "classical" simple recourse structure with overage and underage variables. In fact, it is somewhat similar to the *null recourse* strategy introduced by Powell (1988).

5.4 The Mathematical Formulation

We now formally write down the model that has just been described. In this formulation, we use the definitions and notations for variables and parameters of Section 3 and subsection 5.3 (recall, in particular, that Latin letters refer to deterministic quantities, and Greek letters to stochastic ones).

The First-Stage Problem

Most of the constraints linking the variables of the first-stage problem are identical to those of the deterministic model of Section 3. We recall them for completeness' sake:

Identified Customer Demand.

$$\sum_{t' \le t} \sum_{j \in J_i^{t'}} v_{ji}^{t'} = X_i^t - K_i^t \quad \text{for all} \quad i \in I^t,\ t = 1,2,...,T. \tag{11}$$

Identified Customer Supply.

$$\sum_{j \in J_s^t} v_{sj}^t = Y_s^t \quad \text{for all} \quad s \in S^t,\ \ t = 1,2,...,T. \tag{12}$$

Stocks at Non-Port Depots in Period 1.

$$w_{jj}^1 = w_{jj}^0 + b_j^1 + \sum_{k \in D \cup H \mid \tau_{kj} = 0} w_{kj}^1 + K_j^1 - \sum_{k \in D \cup H} w_{jk}^1 - \sum_{i \in I_j^1} v_{ji}^1 \tag{13}$$

for all $j \in D$.

Port Supply, Demand and Stock in Period 1.

$$w_{jj}^1 = w_{jj}^0 + b_j^1 + \sum_{k \in D \cup H \mid \tau_{kj} = 0} w_{kj}^1 + K_j^1 - \sum_{k \in D \cup H} w_{jk}^1 - \sum_{i \in I_j^1} v_{ji}^1 - X_j^1 + Y_j^1 + e_j^1 \tag{14}$$

for all $j \in H$.

Balancing Flows.

$$l_{jk}^t \le w_{jk}^t \le u_{jk}^t \quad \text{for all } j \in D \cup H,\ k \in D \cup H,\ t = 1,2,...,T\ . \tag{15}$$

To these constraints, we add a set of new equations which allow for the computation of the auxiliary variables $\tilde{w}_j^t$:

Net Inflow (Outflow) from Deterministic Decisions at Depots in Periods 2 to T.

$$\tilde{w}_j^t = \sum_{t' \le t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{t'} + \sum_{s \in S_j^{t'}} v_{sj}^{t'} \right] + K_j^t - \sum_{k \in D \cup H} w_{jk}^t - \sum_{i \in I_j^t} v_{ji}^t \tag{16}$$

for all $j \in D \cup H,\ t = 1,2,...,T$.

Also, all variables are restricted to be non-negative, with the exception of the $\tilde{w}_j^t$'s which are unrestricted in sign.

The objective function of the first-stage model is now to minimize the sum of the costs associated with the deterministic variables, plus the recourse function of the second-stage problem:

$$\sum_{t = 1,2,...,T} \sum_{j \in D \cup H} \left[\sum_{i \in I_j^t} c_{ji}^t\, v_{ji}^t + \sum_{k \in D \cup H} c_{jk}^t\, w_{jk}^t + \sum_{s \in S^t} \sum_{j \in J_s^t} c_{sj}^t\, v_{sj}^t \right] \tag{17}$$

$$+ \sum_{j \in D \cup H} \left[c_j^1\, w_{jj}^1 + \bar{c}_j^1\, b_j^1 \right] + \sum_{j \in H} \bar{c}_j^1\, e_j^1$$

$$+ \sum_{j \in D \cup H} Q_j(w_{jj}^1, \underline{\tilde{w}}_j)$$

The Second-Stage Problem

As was already pointed out at the end of the previous subsection, the second-stage problem is separable by depot. We thus write a second-stage subproblem for each depot.

For a given depot j, $j \in D \cup H$, the second-stage variables are the stocks of empty containers at the end of each period (ω_j^t), and the number of containers borrowed in each period (β_j^t), for periods $t = 2,...,T$. Aside from non-negativity restrictions, there is a single set of constraints linking these variables:

Depot Inventory Balance in Periods 2 to T.

$$\omega_j^t = \omega_j^{(t-1)} + \tilde{w}_j^t + \beta_j^t - \Lambda_j^t \quad \text{for all } t = 2,...,T, \tag{18}$$

where $\omega_j^1 \equiv w_{jj}^1$ by convention.

The objective function of the second-stage subproblem for depot j, given values for w_{jj}^1 and $\underline{\tilde{w}}_j$, is to minimize the expected costs of holding and borrowing containers:

$$Q_j(w_{jj}^1, \underline{\tilde{w}}_j) = E_{\underline{\Lambda}} \left\{ \underset{\underline{\omega}_j, \underline{\beta}_j}{\textit{minimum}} \left[\sum_{t=2,...,T} (c_j^t \omega_j^t + \vec{c}_j^t \beta_j^t) \,|\, \underline{\Lambda}_j \right] \right\} \tag{19}$$

where, $\underline{\Lambda}_j$, $\underline{\omega}_j$ and $\underline{\beta}_j$ represent the vectors of random net demands (Λ_j^t), inventory levels (ω_j^t) and borrowing levels (β_j^t), respectively, for $t = 2,...,T$.

If we assume a reasonable cost structure for the problem (for instance, non-increasing borrowing costs), it is never advantageous to borrow containers in a period while maintaining a positive inventory at he end of the same period. This allows us to determine the optimal values of the second-stage variables, for any given realization of the random parameters, through a set of recursive relations:

$$\hat{\omega}_j^t(w_{jj}^1, \underline{\tilde{w}}_j \,|\, \underline{\Lambda}_j) = [\, \hat{\omega}_j^{(t-1)}(w_{jj}^1, \underline{\tilde{w}}_j \,|\, \underline{\Lambda}_j) + \tilde{w}_j^t - \Lambda_j^t \,]^+ \quad \text{,for all } t = 2,...,T, \tag{20}$$

$$\hat{\beta}_j^t(w_{jj}^1, \underline{\tilde{w}}_j \,|\, \underline{\Lambda}_j) = [\, \Lambda_j^t - \hat{\omega}_j^{(t-1)}(w_{jj}^1, \underline{\tilde{w}}_j \,|\, \underline{\Lambda}_j) - \tilde{w}_j^t \,]^+ \quad \text{,for all } t = 2,...,T, \tag{21}$$

where $[x]^+$ denotes the positive part of x (i.e., max $(0, x)$), and $\hat{\omega}_j^1(w_{jj}^1, \underline{\tilde{w}}_j \,|\, \underline{\Lambda}_j) \equiv w_{jj}^1$.

It is easily verified that the depot recourse functions and, consequently, the global recourse function of the problem, are convex and piecewise linear, for discrete $\underline{\Lambda}_j$ variables.

To solve the model, several approaches could be considered. However, keeping in mind that in most typical applications the size of the network is large (even after performing a substantial aggregation of demands and supplies), one should focus on techniques which exploit the special structure of the problem. In our context, stochastic quasigradient methods, which solve a sequence of subproblems associated with specific realizations of the random parameters (see Ermoliev, 1988, for further details), seem particularly attractive since they allow us to exploit the network structure of the problem (each subproblem is a minimum cost flow problem). We intend to further explore this avenue in the near future.

6. COMPARISON WITH OTHER FLEET MANAGEMENT PROBLEMS

The empty container allocation problem has received up to now very little attention in the literature, and, to our knowledge, there does not exist any model for that specific problem to which the models described in Sections 3, 4 and 5 could be meaningfully compared. On the other hand, operational fleet management problems arising in other contexts (railroad freight transportation and truckload trucking) have been more extensively studied and a number of models have been developed for them (see Dejax and Crainic, 1987, and Powell, 1988, for detailed reviews on this topic). The purpose of this section is to examine and discuss the similarities and differences between our models and some of those.

The *stochastic empty railcar distribution* problem is probably the one which most closely resembles the empty container allocation problem. Apart from the fact that it is usually formulated for a single type of railcars at a time (it is thus a single-commodity problem), one has the same general structure of vehicles (railcars in this instance) becoming free which must be distributed to locations (yards) where there is a demand for empties. Among the models that were proposed for this problem, let us mention those of Mendiratta and Turnquist (1982), Jordan and Turnquist (1983), and Beaujon and Turnquist (1991).

The *dynamic vehicle allocation (DVA)* problem has been defined and studied by Powell, Sheffi and Thiriez (1984), Powell (1986, 1987, 1988), Frantzeskakis and Powell (1990), and Powell and Frantzeskakis (1992), in the context of truckload trucking. In this problem, one has to allocate available trucks to potential loads over a time-space network. Its general structure is thus similar to that of the container allocation and the railcar distribution problems, but several elements (most notably, the very high level of uncertainty regarding future demands and the range of options open to the decision-maker) give to the DVA a flavor of its own.

Let us now examine the main aspects on which our container allocation models differ from the aforementioned ones.

Multi-commodity vs single-commodity

The existence of several container types and of complex substitution rules is certainly one of the factors which most distinguishes the container allocation problem from dynamic vehicle allocation problems for other transportation modes. As a consequence, the notion of

equivalent container had to be introduced into the multicommodity formulation.

Empty only vs combined empty and loaded movements

With regards to the allocation process itself, only empty movements appear as decision variables in the container allocation models (however, as indicated in Section 1, loaded movements are accounted for in the overall planning system and in the routing model), while in many of the previous models (Powell, 1986, 1987, 1988; Beaujon and Turnquist, 1991) both empty and loaded movements are considered. This stems from the fact that loaded containers are shipped outside of the land network (they effectively "disappear" for all planning purposes), while in the other cases the loaded vehicles remain inside the boundaries of the network to become available again for allocation within a short span of time. The same vehicle may thus have to be allocated several times within the planning horizon, which is practically never the case in the land container distribution context. An unexpected consequence of this fact is that current allocation decisions have a lesser impact on the future state of the system under consideration. It is hoped that this may help reduce the feedback distortions caused by the limited planning horizon, thus allowing us to work with a shorter horizon.

Demand handling

In the container allocation problem, all customer demand must absolutely be met within specified delivery windows. In the other models, demands occur at specific instants in time. Requests which cannot be immediately satisfied are either lost (Mendiratta and Turnquist, 1982; Powell, 1986, 1987, 1988) or backordered with a unit penalty per period of delay (Jordan and Turnquist, 1983; Beaujon and Turnquist, 1991). Note also that, in some of the DVA models, it is possible to turn down globally unprofitable loads, even if trucks are available to move them. Consequently, the empty container allocation models are stated as cost minimization problems, while all models which allow for demand to be lost or turned down must be formulated as profit maximization problems.

It is also important to remark that, in models for other transportation modes, the exact location of demand requests is not specified: these occur at railyards or in cities which are permanent nodes in the network structure. Hence, the use of specific location of identified customer requests, and of a variable network structure within the allocation process itself, is new.

Particular Characteristics

Under this label, we include the rules constraining the balancing flows, the external demands and supplies of empties, the relationships with partner shipping companies, which obviously are all specific features of the container allocation problem, not to be found in any of the other problems.

Some of these features had an important impact on the formulations proposed. In the case of the stochastic single commodity problem for instance, the possibility of bringing in containers from outside the system (e.g. by borrowing from partner companies) allowed us to impose deterministic depot-to-depot flows in all periods, even though empty container inventories are random variables. This is in stark contrast with the models proposed by Jordan and Turnquist (1983) and Powell (1986). In these models, which are also based on the *null recourse* assumption of holding in stock vehicles that are not required in a given period, the stochastic inventories lead to stochastic flows between depots. This is because, without the possibility of introducing new vehicles in the system, one cannot make firm decisions on future movements, since these could turn out to be infeasible. In such a situation, the best that one can do at the beginning of the planning horizon is to allocate fractions of the available inventories to specific movements, which implies stochastic flows.

In summary, it can be said that while the empty container allocation problem presents a general structure which is similar to that of other operational fleet management problems, the models described in this paper differ substantially from previous models proposed for these problems.

7. IMPLEMENTATION ISSUES

Although the main objective of this paper is the presentation of formulations of empty container allocation models, several issues related to their implementation and use in actual practice are relevant and are discussed below. We do not aim to be exhaustive. We think, however, that presenting and discussing these items will enhance the understanding of the modeling framework and will put our research effort into its proper perspective.

7.1 Modeling context

Several extensions may be envisioned to our modeling framework. Some of these are rather obvious, even though of significant interest for actual applications. Minimum inventory levels at depots and contractual obligations to send empty containers from customer sites back to the original port, are two examples of this case. Some others, such as nonlinearities in the objective function, are more theoretically interesting, even though in the current state of the available data no practical application may yet be contemplated.

Our interest in this section does not lie, however, with elaborating on this type of model extensions. We are rather concerned with the organizational environment where our models may be applied and the variants thereof. It is clear that the present framework corresponds to a centralized organization, where decisions concerning the allocation of empty containers over the entire territory are taken by a unique authority which considers the network as a whole. This is indeed a widespread organizational scheme. Alternate types of organization are, however, equally possible.

One such alternative corresponds to a "bilevel" structure, made up of a central authority and of several regional ones. Here, briefly, the territory is divided into several regions, each of which is managed by its own authority. The latter controls the operations inside its region, as well as the movements with origin inside the region but with destinations outside its borders. The central authority performs the strategic/tactical planning and coordinates the interregional exchanges, especially the massive balancing flows and the relationships with the outside world.

The modeling framework we have presented offers an adequate representation for this case. The variant for a regional authority is a straightforward application, where balancing

movements are treated somewhat similarly to port imports and exports. The central authority model is principally characterized by its representation of the network: the entry and exit points of the system are emphasized, as well as the interdepot balancing arcs, while the regional subnetworks are suitably aggregated. Particular care has to be taken concerning the proper representation of supplies and demands at the nodes of this network, but the present allocation modeling framework extends without much effort to the "bilevel" case.

Other types of organization may be imagined and may present a research interest. For example, one may contemplate the case where each regional authority operates nearly autonomously, while the central authority is only concerned with the relationships with the external world. A possible modeling avenue would be to treat the demand/supply balancing problem as an *n*-person cooperative game. Yet, as we are not aware at the present time of any practical applications of such a model, we do not plan to further explore this alternative in the near future.

7.2 Data requirements

An empty container allocation model, similarly to any other operational model of a large and complex organization, requires for its proper use in an application environment that abundant data be readily available. This information is varied in nature and may be either used directly (e.g., demand figures) or indirectly (e.g, previous routing decisions and their cost implications) to infer correct approximate figures and formulas, and to update them every time the horizon is advanced. Thus, the models have been designed as part of a decision support system for the short term land transportation planning of loaded and empty containers (see Figure 1), and their appropriate implementation assumes the existence of a reliable information system. To discuss the definition, origin, evaluation/approximation and update procedures for all data items required by an allocation model is clearly beyond the scope of this paper. Hereafter, we only mention some of the most important issues related to specifying demands, supplies and costs.

Accurate and constantly updated forecasts of the supply and demand of empty containers are critical for a successful implementation of an allocation model. Preliminary studies (Gendreau et al., 1991) have shown that simple forecasting methods cannot capture the complexity of the demand mechanism at work and similar conclusions could probably be drawn with

regards to the supplies. A more careful analysis reveals that observed demands and supplies are indeed the result of the superposition of various processes with widely different characteristics. To obtain good forecasts, it is thus necessary to identify as clearly as possible these processes, to monitor them closely through the management information system of the company, to derive good individual forecasts by process, and finally to merge all these forecasts into specific probability distributions.

We use linear cost functions for all activities considered in the allocation models: holding, substitution, borrowing, transport, etc. This is a reasonable hypothesis for an operational model, since most cost nonlinearities, such as potential economies of scale offered by certain transportation modes, have been considered by the strategic/tactical model. Besides, it reflects current industrial practice and data availability. Note also that *handling* costs, corresponding to the various operations and intermodal transshipment of containers at ports, depots and customer sites do not appear in our formulations since the planning of these operations is not part of the allocation decision process.

Container *holding* and *substitution* costs at depots are generally well known. On the other hand, even if *transportation* costs represent the most significant part of the total system costs, one has a less secure knowledge of these, since maritime container shipping companies do not usually own land transportation means, and must therefore contract the services of carriers of various modes. In the case of trucking, they use the services of a large number of local companies, and the cost of transportation is often determined on a case by case basis, according to many factors such as the geographical area, the technical association of truck size and type and container types, the distance to travel, the insurance of return freight or paid return, etc. For rail shipments, they call for the services of national railway companies, which propose an equally complex tariff structure, where the cost of transportation depends on the origin and destination zones, the type, size, weight and number of containers to be shipped together, etc.

We have initiated an investigation aimed at determining formulae for modal unit transportation costs, which take into account these particularities, and which are based on historical data. The initial results are extremely encouraging (Vasseur, 1989; Crainic, Dejax and Gendreau, 1989). However, for any given origin-destination pair, the particular transportation arrangement may vary from period to period, especially when trucking is involved, and this

routing information may not be available at allocation time. Therefore, we plan to integrate a cost update mechanism to the decision support system, to re-estimate the unit costs prior to the advancement of the planning horizon, based on the actual routing decisions.

7.3 Ending the planning horizon

The models of the previous sections cover only a planning horizon of T periods, while the operations of the company will extend in the future for a much longer span of time. This situation, which is encountered in almost all dynamic models, requires careful scrutiny since one would like to minimize (hopefully nullify) the distortions created by the use of a finite horizon model to represent an, for all practical purposes, infinite horizon reality.

Obviously, using the models in a rolling horizon mode helps to mitigate the negative effects of this situation, but it does not by itself eliminate them. It is well-known that, for certain types of models, end-of-horizon distortions will tend to feed back on the values of the decision variables, even in the first period (see, for instance, Hughes and Powell 1988, for a more detailed discussion on this topic). It is thus very important to carefully select the length of the planning horizon and the *end-of-horizon conditions*, in order to obtain valid solutions.

With regards to the length of the planning horizon, two key factors have to be taken into account. One must first be aware that, for any real-life application of the models, the number of decision variables in any period will already be fairly large. This implies that the length of the planning horizon must be limited to some reasonable value (say between 10 and 20 periods), if the models are to be computationally tractable.

The second factor to be considered is the information available on the future supply and demand of empty containers. As we go down the time axis, we move from an area where a large portion of the supply and demand requests can be deemed as certain, to one where everything (or almost) is uncertain. Ideally, the length of the planning horizon should be such as to cover certain demands and supplies (identified customer requests and known import and export movements).

If such an ideal planning horizon is possible, one may then specify fairly simple end-of-horizon conditions to prevent unrealistic solutions. These conditions should be such as to force reasonable values for the empty container stocks in each depot at the end of period T.

(These values may be derived from historical experience and from practical operational constraints, or from an empty container inventory model.) In deterministic models, this can be accomplished by imposing lower and, eventually, upper bounds on these stocks:

$$l_j^T \leq w_{jj}^T \leq u_j^T \quad \text{for all } j \in D \cup H, p \in P.$$

One could also include a *salvage value* term in the holding cost functions for period T, to account for the marginal value of the various container types at each depot (Hughes and Powell, 1988). This last option could also be implemented in the stochastic formulation. Obviously, in any practical application, some experimentation with possible values for these bounds and salvage terms would be necessary to ascertain that the chosen values are indeed correct.

When no feasible ideal horizon exists, the specification of the end-of-horizon conditions becomes a much more difficult task. In such a situation, one would first have to deal with "split delivery windows", i.e., delivery windows which fall partly within the planning horizon and partly outside of it. Deciding whether or not the corresponding demands should be included in the planning horizon is not obvious. In most cases, one would probably be well-advised to exclude them and to slightly adjust the lower bounds on the empty container stocks at the depots from which these demands should be satisfied later on (this is known from the strategic/tactical plan). Similar remarks apply to known import supplies and export demands occurring after the last period of the horizon.

In summary, it must be stressed that there is no cut and dry answer to the problems created by the necessity to work with a limited planning horizon. The modeling framework offers a number of means (bounds, salvage values) which can be used to tackle these problems, but the specific solution to implement in any practical application of the model is highly dependent upon the situation at hand.

CONCLUSIONS

We have described the problem of allocating empty containers in a land distribution and transportation system, and have identified its basic structure and main characteristics. We have introduced two dynamic deterministic formulations for the single and multicommodity cases, respectively, which, although based on actual experience with a large container shipping company, offer a general modeling framework for this class of problems. This framework reflects the operational and planning complexity of the land component of the international maritime shipping of containers, particularly the space and time dependency of events, as well as specific operational characteristics, such as substitutions, relationships with partner companies, imports and exports, massive equilibration flows, etc. Finally, we provided a mathematical formulation for handling, in the single commodity case, the uncertainty of demand and supply data, characteristic of container allocation and distribution problems.

To our knowledge, these are the first models specifically aimed at the empty container land allocation problem. We have attempted, therefore, to explore as comprehensively as possible the various modeling issues related to this problem, while proposing formulations that are computationally tractable. This has allowed us to highlight the originality of the problem and to explain the differences which may be observed between our modeling framework and formulations for similar allocation problems arising in the context of other transportation modes.

We have sketched some of the most promising avenues for the development of efficient solution techniques for the proposed models. This, together with a continuing exploration of richer stochastic formulations, especially for the multicommodity case, constitutes the next stage of our research program. We also plan to continue the investigation of the loaded and empty container routing problem and that of the operational planning system required for an efficient actual implementation of the models.

ACKNOWLEDGEMENTS

The authors wish to thank Professor Warren B. Powell, Guest Editor of this Special Issue, the Associate Editor and the referees for their thorough review of the manuscript; their suggestions and comments have helped us write a better paper. The authors are also grateful to the Canadian *Natural Sciences and Engineering Research Council*, to the *Fonds F.C.A.R.* of the Province of Québec, and to the *Programme de coopération France – Québec* , who have provided the financial support for this research.

REFERENCES

Antonisse K. 1988, A Decision Support System for Container Replenishment, TIMS XXVIII/EURO IX Joint Meeting, Paris, July 6-8.

Antush R.M. 1989, Research Issues at Sealand for the 1990's, ORSA/TIMS Joint National Meeting, New York, October 16-18.

Beaujon G.J. and Turnquist M.A. 1991, A Model for Fleet Sizing and Vehicle Allocation, *Transportation Science* 25(1), 19-45.

Crainic T.G., Dejax P.J. and Delorme L. 1989, Models for Multimode Multicommodity Location Problems with Interdepot Balancing Requirements, *Annals of Operations Research* 18, 279-302.

Crainic T.G., Dejax P.J. and Gendreau M. 1989, Transport terrestre multimodal de conteneurs: problématique et modèles, in *Logistique/Logistics: Production, Distribution, Transport / Manufacturing, Distribution, Transportation,* Proceedings of the Joint *Conference on the Practice and Theory of Operations Management* and *4es journées francophones sur la Logistique et les Transports*, 105-113, AFCET, Paris.

Crainic T.G., Gendreau M. and Dejax P.J. 1990, Modelling the Container Fleet Management Problem Using a Stochastic Dynamic Approach, *Operational Research'90* (H.E. Bradley Editor), 473-486, Pergamon Press.

Crainic T.G., Gendreau M. and Dejax P.J. 1991, *A Dynamic Stochastic Model for the Allocation of Empty Containers*, publication #713, Centre de recherche sur les transports, Université de Montréal.

Dejax P.J., Benamar F., Crainic T.G. and Gendreau M. 1992, Short Term Container Fleet Management: Issues, Models and Tools, Proceedings of *The 6th World Conference on Transport Research*, Lyon, France, June 29 - July 3, 1992.

Dejax P.J. and Crainic T.G. 1987, A Review of Empty Flows and Fleet Management Models in Freight Transportation, *Transportation Science* 21(4), 227-247.

Dejax P.J., Crainic T.G. and Delorme L. 1988, Strategic/Tactical Planning of Container Land Transportation Systems, TIMS XXVIII/EURO IX Joint Meeting, Paris, July 6-8.

Ermoliev Y. 1988, Stochastic Quasigradient Methods, in *Numerical Techniques for Stochastic Optimization* (Y. Ermoliev and R. Wets, Eds.), Springer Verlag, Berlin.

Hughes R.E. and Powell W.B. 1988, Mitigating End Effects in the Dynamic Vehicle Allocation Model, *Management Science* 34(7), 859-879.

Frantzeskakis L.F. and Powell W.B. 1989, A Successive Linear Approximation Procedure for Stochastic, Dynamic Vehicle Allocation Problems, *Transportation Science* 24(1), 40-57.

Gendreau M., Crainic T.G., Dejax P. and Steffan H. 1991, Forecasting Short-Term Demand for Empty Containers: A Case Study, *Transportation Research Record* 1328, 98-103..

Jarke M. 1982, Developing Decision Support Systems: A Container Management Example, *Int. J. Policy Anal. Inform. Syst.* 6, 351-372.

Jordan W.C. and Turnquist M.A. 1983, A Stochastic, Dynamic Network Model for Railroad Car Distribution, *Transportation Science* 17, 123-145.

Mendiratta V.B. and Turnquist M.A. 1982, A Model for the Management of Empty Freight Cars, *Transportation Research Record* 838, 50-55.

Powell W.B. 1986, A Stochastic Model of the Dynamic Vehicle Allocation Problem, *Transportation Science* 20(2), 117-129.

Powell W.B. 1987, An Operational Planning Model for the Dynamic Vehicle Allocation Problem with Uncertain Demands, *Transportation Research* 21B, 217-232

Powell W.B. 1988, A Comparative Review of Alternative Algorithms for the Dynamic Vehicle Allocation Problem, in: *Vehicle Routing : Methods and Studies* (B.L. Golden and A.A. Assad, Editors), pp. 249-291, North-Holland, Amsterdam.

Powell W.B., Sheffi Y. and Thiriez S. 1984, The Dynamic Vehicle Allocation Problem with Uncertain Demands, in *9th Int. Symp. on Transportation and Traffic Theory* (J. Volmuller and R. Hamerslag, Editors), VNU Science Press.

Powell W.B. and Frantzeskakis L.F. 1992, Restricted Recourse Strategies for Stochastic, Dynamic Networks, *Transportation Science* (to appear).

Rockafellar T.R. and Wets R.J. 1991, Scenarios and Policy Aggregation in Optimization Under Uncertainty, *Mathematics of Operations Research* 16(1), 119-147.

Vasseur C. 1989, Les coûts du transport ferroviaire par conteneurs en France: étude de cas, Publication # 650, Centre de recherche sur les transports, Université de Montréal.

Wets R. 1974, Stochastic Programs with Fixed Resources: The Equivalent Deterministic Problem, *SIAM Review* 16, 309-339.